

FD - 610111, 121 11271000
 PD - 12/01/99
 SK 601/68229

GenCore version 5.1.6
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
 Run on: June 18, 2003, 05:32:01 ; Search time 3098 Seconds
 (without alignments)
 10342.863 Million cell updates/sec
 Title: US-09-874-062-2
 Perfect score: 1101
 Sequence: 1 ctgggaaccaccacccagacc.....caaaaaaaaaaaaaaaaaaaaaa 1101

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
 Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database :	GenEmbl:*
1:	gb_ba:*
2:	gb_hgtg:*
3:	gb_in:*
4:	gb_om:*
5:	gb_ov:*
6:	gb_pat:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sts:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vl:*
15:	em_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_or:*
22:	em_ov:*
23:	em_pat:*
24:	em_ph:*
25:	em_pi:*
26:	em_ro:*
27:	em_sts:*
28:	em_un:*
29:	em_vl:*
30:	em_hgtg_hum:*
31:	em_hgtg_inv:*
32:	em_hgtg_other:*
33:	em_hgtg_mus:*
34:	em_hgtg_pln:*
35:	em_hgtg_rod:*
36:	em_hgtg_man:*
37:	em_hgtg_vrt:*
38:	em_sy:*
39:	em_hgtg_hum:*
40:	em_hgtg_mus:*
41:	em_hgtg_other:*

score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	909.2	82.6	347572	2	AC107303	AC107303 Homo sapi
2	876.8	79.6	161987	9	AC006070	AC006070 Homo sapi
3	735.6	66.8	174032	2	AC025904	AC025904 Homo sapi
4	669	60.8	671	9	HS406935	AJ406935 Homo sapi
5	403.4	36.6	133392	2	AC099183	AC099183 Rattus no
6	403.4	36.6	182806	2	AC098902	AC098902 Rattus no
7	378.2	34.4	192027	2	AC067715	AC067715 Mus muscu
8	378.2	34.4	200602	10	AL590997	AL590997 Mouse DNA
9	368.2	33.4	197658	2	AL646094	AL646094 Mus muscu
10	368.2	33.4	200602	10	AL590997	AL590997 Mouse DNA
11	356	32.3	174032	2	AC025904	AC025904 Homo sapi
12	356	32.3	347572	2	AC107303	AC107303 Homo sapi
13	354.4	32.2	1099	9	HS406936	AJ406936 Homo sapi
14	354.4	32.2	1203	9	HS406939	AJ406939 Homo sapi
15	354.4	32.2	186078	9	AC037482	AC037482 Homo sapi
16	343.8	31.2	197658	2	AL646094	AL646094 Mus muscu
17	335.4	30.5	1220	9	HS406944	AJ406944 Homo sapi
18	335.4	30.5	186078	9	AC037482	AC037482 Homo sapi
19	335.2	30.4	111637	2	AC100808	AC100808 Homo sapi
20	328.4	29.8	1093	9	HS406943	AJ406943 Homo sapi
21	324.2	29.4	111637	2	AC100808	AC100808 Homo sapi
22	319	29.0	1080	9	BC004180	BC004180 Homo sapi
23	319	29.0	1080	9	BC004212	BC004212 Homo sapi
24	314	28.5	55568	2	AC015850_3	AC015850 Homo sapi
25	312.6	28.4	182806	2	AC098902	AC098902 Rattus no
26	310.4	28.2	118444	10	AL590992	AL590992 Mouse DNA
27	303.2	27.5	924	9	HS4296168	AJ296168 Homo sapi
28	297.4	27.0	942	9	HS406937	AJ406937 Homo sapi
29	297	27.0	110000	2	AC015850_0	AC015850 Homo sapi
30	287	26.1	110000	2	AC015850_0	AC015850 Homo sapi
31	270.8	24.6	1064	9	HS406941	AJ406941 Homo sapi
32	269.2	24.5	1037	9	HS406945	AJ406945 Homo sapi
33	265.6	24.1	485	4	OAHKP	X73462 O.aries mRN
34	263.2	23.9	133392	2	AC099183	AC099183 Rattus no
35	252	22.9	3489	6	AR072952	AR072952 Sequence
36	252	22.9	3489	6	AX107938	AX107938 Sequence
37	252	22.9	3489	14	KSU52064	U52064 Kaposi's sa
38	252	22.9	32207	6	AR065852	AR065852 Sequence
39	252	22.9	32207	6	AR127850	AR127850 Sequence
40	252	22.9	32207	6	AR194752	AR194752 Sequence
41	252	22.9	137508	14	KSU75698	U75698 Kaposi's sa
42	251.4	22.8	110000	2	AC015850_2	Continuation (3 of
43	250.8	22.8	782	9	HS406934	AJ406934 Homo sapi
44	241.2	21.9	63280	2	AC105960	AC105960 Mus muscu
45	238.2	21.6	28559	14	AF148805	AF148805 Kaposi's

ALIGNMENTS

RESULT 1	AC107303	347572 bp	DNA	linear	HTG 09-MAY-2002
AC107303	Homo sapiens chromosome 3 clone RP11-595122, WORKING DRAFT				
LOCUS	SEQUENCE, 32 unordered pieces.				
DEFINITION	AC107303				
ACCESSION	AC107303.3	GI:20335965			
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT.				
KEYWORDS	Homo sapiens				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,				
	Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,				

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowles, M., Brown, E., Brown, M., Bryant, N.P.,
Buay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Haves, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, J., Johnson, R., Jollivet, S., Joudah, S.,
Karissom, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L., Oulles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Soneike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vaquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Worley, K.,
Williams, G., Williamson, A., Wleczka, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 347572)
Worley, K.C.

Direct Submission
Submitted (18-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 347572)
Worley, K.C.

Direct Submission
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:18449841.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HDSS
Center clone name: RP11-595122
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 331666 bases at least Q40
Consensus quality: 339305 bases at least Q30
Consensus quality: 344140 bases at least Q20
Estimated insert size: 172965; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

COMMENT

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
consists of 32 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 2351: contig of 2351 bp in length
2352 2451: gap of unknown length
2452 6192: contig of 3741 bp in length
6193 6292: gap of unknown length
6293 10339: contig of 4047 bp in length
10340 10439: gap of unknown length
10440 13681: contig of 3242 bp in length
13682 13781: gap of unknown length
13782 18293: contig of 4512 bp in length
18294 18393: gap of unknown length
18394 22935: contig of 4541 bp in length
22936 23034: gap of unknown length
23035 26605: contig of 3571 bp in length
26606 26705: gap of unknown length
26706 31631: contig of 4925 bp in length
31632 31730: gap of unknown length
31731 36997: contig of 5267 bp in length
36998 37097: gap of unknown length
37098 41631: contig of 4534 bp in length
41632 41731: gap of unknown length
41732 48452: contig of 6721 bp in length
48453 48552: gap of unknown length
48553 55906: contig of 7254 bp in length
55907 55906: gap of unknown length
55907 65188: contig of 9282 bp in length
65189 65288: gap of unknown length
65289 71857: contig of 6569 bp in length
71858 71957: gap of unknown length
71958 80512: contig of 8555 bp in length
80513 87649: contig of 7037 bp in length
87650 87749: gap of unknown length
87750 94122: contig of 6373 bp in length
94123 94222: gap of unknown length
94223 103065: contig of 8842 bp in length
103066 103164: gap of unknown length
103165 112059: contig of 8895 bp in length
112060 112159: gap of unknown length
112160 120884: contig of 8725 bp in length
120885 120984: gap of unknown length
120985 134158: contig of 13174 bp in length
134159 134258: gap of unknown length
134259 146831: contig of 12573 bp in length
146832 146931: gap of unknown length
146932 158213: contig of 11282 bp in length
158214 158313: gap of unknown length
158314 172201: contig of 13888 bp in length
172202 172301: gap of unknown length
172302 184228: contig of 11927 bp in length
184229 184328: gap of unknown length
184329 195498: contig of 11170 bp in length
195499 195598: gap of unknown length
195599 210746: contig of 15148 bp in length
210747 210846: gap of unknown length
210847 226328: contig of 15482 bp in length
226329 226429: gap of unknown length
226430 246181: contig of 19753 bp in length
246182 246281: gap of unknown length
246282 265511: contig of 19229 bp in length
265512 265611: gap of unknown length
265611 293687: contig of 28076 bp in length
293688 347572: gap of unknown length
347573 347572: contig of 53786 bp in length.

Location/Qualifiers
1. 347572

FEATURES
source

SEQ ID NO: 2, Result 2
Database: GenBank; AC NO: AC006070

/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="3"	
/clone="RP11-595I22"	
BASE COUNT	107326 a 64859 c 63333 g 108900 t 3154 others
ORIGIN	
Query Match 82.6%; Score 909.2; DB 2; Length 347572;	
Best Local Similarity 91.9%; Pred. No. 6e-148;	
Matches 1002; Conservative 0; Mismatches 13; Indels 75; Gaps 1;	
QY	1 CTGGGAACCCAGAGCTGCTCAAGCTCTCGCCAGCAGAGCTGCTGCCGCCGCCAGC 120
Db	105475 CTGGGAACCCAGAGAGCTCAAGCTCTGAGCGCATGGTCAGCTCTGCTGGCTCT 60
QY	61 GTCTGCTGTGACAGAGCTGCTCAAGCTCTCGCCAGCAGAGCTGCTGCCGCCGCCAGC 120
Db	105535 GTCTGCTGTGACAGAGCTGCTCAAGCTCTCGCCAGCAGAGCTGCTGCCGCCGCCAGC 105594
QY	121 TGTGCGCAGACCACTGCTGCGAGGACCACTGCTGCCGCCGCCAGCTGCTGCAATTTCCAGT 180
Db	105595 TGTGCGCAGACCACTGCTGCGAGGACCACTGCTGCCGCCGCCAGCTGCTGCAATTTCCAGT 105654
QY	181 TGTGCGAGCCCTTCTGCTGTATCTCCAGCTGCTGCAAAACCACTGCTGCGCTGACCACT 240
Db	105655 CGCTGAGGCCCTTCTGCTGTATCTCCAGCTGCTGCAAAACCACTGCTGCGCTGACCACT 105714
QY	241 TGTGCGAGCAGCAGCTGCTGCCGCCGCCAGCTGCTGCAATTTCCAGTTCAGGCTTCC 300
Db	105715 TGTGCGAGCAGCAGCTGCTGCCGCCGCCAGCTGCTGCAATTTCCAGTTCAGGCTTCC 105774
QY	301 TGTGCTATCTCCAGCTGCTGCAAAACCACTGCTGCGAGCAGCACTGCTGCGGCCGCCAGC 360
Db	105775 TGTGCTATCTCCAGCTGCTGCAAAACCACTGCTGCGAGCAGCACTGCTGCGGCCGCCAGC 105818
QY	361 TGTGCTATCTCCAGCTGCTGCGAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCACT 420
Db	105819 -----C----- 105819
QY	421 TGTGCGAGCAGCAGCTGCTGCCGCCGCCAGCTGCTGCAATTTCTAGTTCAGGCTTCC 480
Db	105820 TGTGCGAGCAGCAGCTGCTGCCGCCGCCAGCTGCTGCAATTTCTAGTTCAGGCTTCC 105879
QY	481 TGTGCTATCTAGCTGTTGCAAAACCACTGCTGCGCAGCAGCACTGCTGCGGCCGCCAGC 540
Db	105880 TGTGCTATCTAGCTGCTGCAAAACCACTGCTGCGCAGCAGCACTGCTGCGGCCGCCAGC 105939
QY	541 TGTGCTATCTCCAGCTGCTACAGGCCGCCAGTGTGCGCAGGCTGCTGCTGCCGCCGCCGCT 600
Db	105940 TGTGCTATCTCCAGCTGCTACAGGCCGCCAGTGTGCGCAGGCTGCTGCTGCCGCCGCCGCT 105999
QY	601 TGTGCTATCTAGTTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db	106000 TGTGCTATCTAGTTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 106059
QY	661 AGTGTGCCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db	106060 AGTGTGCCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 106119
QY	721 TGTGCTGAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db	106120 TGTGCTGAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 106179
QY	781 GCACAGATGATCTATTACAGAGACATGCTGGACTTCTGATGCTGCTGCTGCTGCTGCTGCTGCT 840
Db	106180 GCACAGATGATCTATTACAGAGACATGCTGGACTTCTGATGCTGCTGCTGCTGCTGCTGCTGCT 106239
QY	841 GACTGATTTGGAAATATTTTATGATGATCTATCTCTCTTTATAGAGTTCATGATTTTATCTCTA 900
Db	106240 GACTGATTTGGAAATATTTTATGATGATCTATCTCTCTTTATAGAGTTCATGATTTTATCTACTA 106299
QY	901 TTGAATCTGAATTTACAGTCAAAATTCACATCAGATGTTTATGAAATCTTTTATCTAAAT 960

Db	106300	TTGAATCTGAATTTACAGTCAAAATTCACATCAGATGTTT	TAGAACTCTTTTATCTTAAT	106359
QY	961	CAATATACATAAATCTTCAAAATGGTATCTCTTAGATGTTTCTCTCTAATGTTTCTG	1020	
Db	106360	CAATATACATAAATCTTCAAAATGGTATCTCTTAGATGTTTCTCTAATGTTTCTG	106419	
QY	1021	GTATCAATTTTTCATGTCGAATTTGTTTGATGTTCCCAATAAACTTCATGTTTCAAA	1080	
Db	106420	GTATCAATTTTTCATGTCGAATTTGTTTGATGTTCCCAATAAACTTCATGTTTCAAA	106479	
QY	1081	GCAAAAAAAA	1090	
Db	106480	GCAACAAAAA	106489	

RESULT 2	AC006070	161987 bp	DNA	linear	PK1 17-DBC-1998
LOCUS	Homo sapiens chromosome 17, clone hRPK.206_C_20, complete sequence.				
DEFINITION	AC006070				
ACCESSION	AC006070.1	GI:4028942			
VERSION	HTG.				
KEYWORDS	Homo sapiens.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 161987)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.				
TITLE	Homo sapiens chromosome 17, clone hRPK.206_C_20				
JOURNAL	Unpublished				
AUTHORS	2 (bases 1 to 161987)				
REFERENCE	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Bontwell, C., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collins, S., Collymore, A., Cooke, P., Corliss, D., Depayre, E., Devon, K., Dewar, K., Donelan, L., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S., Geraghty, K., Grant, G., Hagos, R., Heald, A., Herena, L., Horton, L., Howland, J. C., Jacotot, L., Jones, C., Kuan, L., Karas, A., Lehotzky, J., MacDonald, P., Marquis, N., McEwan, P., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, R., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C., Subramanian, A., Testaye, S., Tichovolsky, N., Tortorella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D., Ye, W. J., Zhao, J. and Zody, M.				

DIRECT SUBMISSION	
JOURNAL	Submitted (26-NOV-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	3 (bases 1 to 161987)
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K., Donelan, L., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galauan, J., Gardyna, S., Grant, G., Hagos, R., Heald, A., Herena, L., Horton, L., Howland, J. C., Jones, C., Kuan, L., Karas, A., Lehotzky, J., MacDonald, P., Marquis, N., McEwan, P., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Testaye, S., Tortorella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
TITLE	Direct Submission
JOURNAL	Submitted (17-DEC-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Dec 17, 1998 this sequence version replaced gi:4006846. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RN/RepeatMasker.html

Only the last 161987 base pairs of this clone are being submitted.

The remainder overlaps accession number AC003958 (WICGR project L295).

FEATURES

source	Location/Qualifiers	
repeat_region	1. .161987	/rpt_family="AluSq"
repeat_region	/organism="Homo sapiens"	16482. .16555
repeat_region	/db_xref="taxon:9606"	/rpt_family="(CAAA)n"
repeat_region	/chromosome="17"	16556. .16727
repeat_region	/map="17"	/rpt_family="AluSq"
repeat_region	/clone="BRPK.206.C.20"	complement(16870. .16897)
repeat_region	complement(29. .218)	/rpt_family="(CAAAAA)n"
repeat_region	/rpt_family="L2"	complement(17088. .17221)
repeat_region	complement(260. .512)	/rpt_family="AluSq/x"
repeat_region	/rpt_family="L2"	complement(17989. .18273)
repeat_region	complement(800. .1013)	/rpt_family="AluSq"
repeat_region	/rpt_family="MER20"	18409. .18438
repeat_region	1140. .1416	/rpt_family="AT_rich"
repeat_region	/rpt_family="AluSq"	complement(18439. .18723)
repeat_region	1419. .1453	/rpt_family="AluSq"
repeat_region	/rpt_family="(CA)n"	19095. .19584
repeat_region	complement(1476. .1515)	/rpt_family="MER31B"
repeat_region	/rpt_family="MIR"	20277. .20562
repeat_region	1589. .1770	/rpt_family="AluSc"
repeat_region	/rpt_family="L1ME3"	22621. .22641
repeat_region	1786. .1905	/rpt_family="AT_rich"
repeat_region	/rpt_family="(TA)n"	23843. .23992
repeat_region	complement(2657. .2784)	/rpt_family="L2"
repeat_region	/rpt_family="MIR"	24184. .24291
repeat_region	3298. .3422	/rpt_family="MIR"
repeat_region	/rpt_family="MIR"	25327. .25351
repeat_region	complement(6000. .6123)	/rpt_family="AT_rich"
repeat_region	/rpt_family="AluSp/a"	complement(25352. .25907)
repeat_region	complement(6127. .6238)	/rpt_family="L1PA2"
repeat_region	/rpt_family="LTR41"	complement(27038. .28106)
repeat_region	6640. .6751	/rpt_family="L1MA4A"
repeat_region	/rpt_family="(CA)n"	complement(28116. .28209)
repeat_region	7330. .7398	/rpt_family="MLT1C"
repeat_region	/rpt_family="MIR"	complement(28515. .28581)
repeat_region	7759. .7899	/rpt_family="purine-rich"
repeat_region	/rpt_family="L1ME2"	complement(28582. .28674)
repeat_region	complement(7909. .8115)	/rpt_family="MLT1C"
repeat_region	/rpt_family="MER58A"	29080. .29354
repeat_region	complement(8116. .8903)	/rpt_family="L1TR16C"
repeat_region	/rpt_family="L1TR1"	29355. .29510
repeat_region	complement(8904. .8928)	/rpt_family="MIR"
repeat_region	/rpt_family="MER58A"	29524. .29569
repeat_region	8935. .9031	/rpt_family="AT_rich"
repeat_region	/rpt_family="L1ME2"	complement(29595. .29863)
repeat_region	9060. .9363	/rpt_family="L1MA9"
repeat_region	/rpt_family="L1ME2"	complement(30929. .31490)
repeat_region	10471. .10518	/rpt_family="L1M3C"
repeat_region	/rpt_family="AT_rich"	32091. .32199
repeat_region	complement(10646. .11003)	/rpt_family="MER81"
repeat_region	/rpt_family="MLT1F"	complement(32754. .33081)
repeat_region	11004. .11209	/rpt_family="L1TR16C"
repeat_region	/rpt_family="L1PA5"	33760. .33915
repeat_region	11218. .11296	/rpt_family="MIR"
repeat_region	/rpt_family="(TA)n"	33929. .33971
repeat_region	11326. .11561	/rpt_family="AT_rich"
repeat_region	/rpt_family="THE1B"	complement(33997. .34259)
repeat_region	complement(11779. .12083)	/rpt_family="L1MA9"
repeat_region	/rpt_family="AluSq"	complement(35330. .35875)
repeat_region	complement(12319. .12588)	/rpt_family="L1M3e"
repeat_region	/rpt_family="MLT1-INTERNAL"	36529. .36637
repeat_region	complement(12942. .13139)	/rpt_family="MER81"
repeat_region	/rpt_family="L2"	complement(37178. .37548)
repeat_region	complement(13361. .13649)	/rpt_family="MLT1B"
repeat_region	/rpt_family="AluSq"	complement(37860. .38117)
repeat_region	complement(14208. .14490)	/rpt_family="polypurine"
repeat_region	/rpt_family="AluY"	complement(38090. .38209)
repeat_region	complement(14636. .14925)	/rpt_family="purine-rich"
repeat_region	/rpt_family="AluSq"	38432. .38695
repeat_region	16356. .16481	/rpt_family="L1TR16C"

Query Match

79.6% Score 876.8 DB 9 Length 161987

Best Local Similarity 94.6%; Pred. No. 2.4e-142; Matches 908; Conservative 0; Mismatches 52; Indels 0; Gaps 0;									
QY	131	CCACCTGCTGCAGGACCACTGCTGCGGCCAGCTGCTGCATTTCCAGTTCTGCTGCAGCG	190						
Db	138927	CCTCTGACCCCATGTCAGCTCCTGCTGTGGCTCTGCTCTGCTCTGACCAAGACTGTGGTC	138986						
QY	191	CTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACCACTGCTGCAGGA	250						
Db	138987	AAGGTCTCGCCAGGAGAGCTGCTGCGGCCAGCTGCTGCCAGACCACTGCTGCAGGA	139046						
QY	251	CCACCTGCTGCGGCCAGCACTGCTGCATTTCCAGTTGCTGCGAGGCTTCTCTGCTGTATCT	310						
Db	139047	CCACCTGCTGCGGCCAGCACTGCTGCATTTCCAGTTGCTGCGAGGCTTCTCTGCTGTATCT	139106						
QY	311	CCAGCTGCTGCAAAACCCAGCTGCTGCAGGACCACTGCTGCGGCCAGCTGCTGCATTT	370						
Db	139107	CCAGCTGCTGCAAAACCCAGCTGCTGCAGGACCACTGCTGCGGCCAGCTGCTGCATTT	139166						
QY	371	CCAGTTGCTGCGAGGCTTCTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGA	430						
Db	139167	CCAGTTGCTGCGAGGCTTCTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGA	139226						
QY	431	CCACCTGCTGCGGCCAGCACTGCTGCATTTCTAGTTGCTGCGAGGCTTCTCTGCTGTATCT	490						
Db	139227	CCACCTGCTGCGGCCAGCTGCTGCATTTCTAGTTGCTGCGAGGCTTCTCTGCTGTATCT	139286						
QY	491	CTAGCTGTTGCAAAACCCAGCTGCTGCCAGCACCTGCTGCGGCCAGCTGCTGTATCT	550						
Db	139287	CTAGCTGTTGCAAAACCCAGCTGCTGCCAGCACCTGCTGCGGCCAGCTGCTGTATCT	139346						
QY	551	CCAGCTGCTACAGGCCCACTGCTGCCAGCCCTCTCTGCTGCGGCCAGCTTCTGCTCATTT	610						
Db	139347	CCAGCTGCTACAGGCCCACTGCTGCCAGCCCTCTCTGCTGCGGCCAGCTTCTGCTCATTT	139406						
QY	611	CTAGTTGCTGTCATCCAGCACTGCTGTGTGTCAGCTGCGCGCTGCCCTTTTCACTGCCCA	670						
Db	139407	CTAGTTGCTGTCATCCAGCACTGCTGTGTGTCAGCTGCGCGCTGCCCTTTTCACTGCCCA	139466						
QY	671	CCACCTGCTGTAGAACCACTGCTTCCACCCCATCTGCTGCGGCCAGCTTCTGCTGCTGAG	730						
Db	139467	CCACCTGCTGTAGAACCACTGCTTCCACCCCATCTGCTGCGGCCAGCTTCTGCTGCTGAG	139526						
QY	731	TGAAGCTGCTCTGGATTCTTGCACTTCTTGCTCAACCTTTCAGTTTCAGGCACACAGTA	790						
Db	139527	TGAAGCTGCTCTGGATTCTTGCACTTCTTGCTCAACCTTTCAGTTTCAGGCACACAGTA	139586						
QY	791	TCTATTTCAGAACATGTGGACTTCTGTATGTCGTGAAACAGAGGCATGGACTGATTTG	850						
Db	139587	TCTATTTCAGAACATGTGGACTTCTGTATGTCGTGAAACAGAGGCATGGACTGATTTG	139646						
QY	851	GAATAATTTTATTAGTATGTTCTCTTTTATAGAGTTTTTATTCCTTATTGAATCTGA	910						
Db	139647	GAATAATTTTATTAGTATGTTCTCTTTTATAGAGTTTTTATTCCTTATTGAATCTGA	139706						
QY	911	ATTACAGTCAAAATCCACATCATGTTTATAGAATTTCTTATCTTAATCAATATACAT	970						
Db	139707	ATTACAGTCAAAATCCACATCATGTTTATAGAATTTCTTATCTTAATCAATATACAT	139766						
QY	971	AAATCTTCAATGGTATCTCTTAGATGTTTCTTCTTAATGTTTCTGCTGATCAATTT	1030						
Db	139767	AAATCTTCAATGGTATCTCTTAGATGTTTCTTCTTAATGTTTCTGCTGATCAATTT	139826						
QY	1031	TCATGTGGAATGTTTGTATGTTCTTCAATAAATCTTCATAGTGTTCAAAGCAAAAAA	1090						
Db	139827	TCATGTGGAATGTTTGTATGTTCTTCAATAAATCTTCATAGTGTTCAAAGCAAAAAA	139886						

RESULT 3

AC025904/c

LOCUS

DEFINITION

Homo sapiens clone RP11-585G19, WORKING DRAFT SEQUENCE, 17
unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC025904
AC025904.2 GI:7382590
HTG: HTGS_PHASE1: HTGS_DRAFT.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174032)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-585G19
Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

2 (bases 1 to 174032)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodgson, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaKocque, K., Lamazares, R., Landers, T., Lehorzky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienda, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodores, J., Tirrell, A., Travers, M., Triguero, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 1, 2000 this sequence version replaced gi:7249428.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBC
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8441
Center clone name: 585_G19
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158771 bases at least Q40
Consensus quality: 167389 bases at least Q30
Consensus quality: 170660 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 172432; sum-of-ctnigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1
* 2588: contig of 2588 bp in length
* 2589: gap of 100 bp
* 2689: contig of 3777 bp in length
* 6466: gap of 100 bp
* 6566: contig of 2281 bp in length

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 671) Rogers, M.A., Langbein, L., Winter, H., Ehmann, C., Korn, B. and Schweizer, J. Characterization of a cluster of human high/ ultrahigh keratin associated proteins on chromosome 17q12-21 Unpublished 2 (bases 1 to 671) Rogers, M.A. Direct Submission Submitted (16-OCT-2000) Rogers M.A., Research Program B, German Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg, Germany 69120, GERMANY Related genomic sequence: AC006070 (139229-139525nt). COMMENT Location/Qualifiers 1. .671 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="17" /map="17q12-21" /clone="sc166" /tissue_type="scalp" 1. .297 /gene="KRTAP4.3" <1. .297 /gene="KRTAP4.3" /function="structural protein" /standard_name="KAP4.3" /codon_start=1 /product="keratin associated protein 4.3" /protein_id="I265544" /db_xref="GI:1265544" /translation="TCCRPSCCISCCRPSCCISCCRPSCCOTTCRPPSCCISCCYR PQCQPSCCRPACCISSCHPSCVSSCRPFSCPTTCRTTCHPFCGSSCC" polyA_signal 157 a 175 c 125 g 214 t BASE COUNT 625. .630 ORIGIN Query Match 60.8%; Score 669; DB 9; Length 671; Best Local Similarity 100.0%; Pred. No. 1.8e-106; Matches 669; Conservative 0; Mismatches 0; Indels 0; Gaps 0; QY 433 ACCTGCTGCCGCCAGCTGCTGATTTCTAGTTGCTGAGGCCCTTCTGCTGATCTCT 492 DB 1 ACCTGCTGCCGCCAGCTGCTGATTTCTAGTTGCTGAGGCCCTTCTGCTGATCTCT 60 QY 493 AGCTGTTGCAAAACCCAGCTGCTGCCAGACCACTGCTGCCGCCGCCAGCTGCTGATCTCC 552 DB 61 AGCTGTTGCAAAACCCAGCTGCTGCCAGACCACTGCTGCCGCCGCCAGCTGCTGATCTCC 120 QY 553 AGCTGCTACAGGCCCCAGCTGCTGCCAGCCCTGCTGCTGCCGCCGCCAGCTGCTGATTTCT 612 DB 121 AGCTGCTACAGGCCCCAGCTGCTGCCAGCCCTGCTGCTGCCGCCGCCAGCTGCTGATTTCT 180 QY 613 AGTTGCTGTATCCAGCTGCTGCTGCCAGCTGCCGCCGCCAGCTGCTGCCGCCGCCAGCTGCTGATTTCT 672 DB 181 AGTTGCTGTATCCAGCTGCTGCTGCCAGCTGCCGCCGCCAGCTGCTGCCGCCGCCAGCTGCTGATTTCT 240 QY 673 ACCTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCCAGCTGCCGCCGCCAGCTGCTGATTTCT 732 DB 241 ACCTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCCAGCTGCCGCCGCCAGCTGCTGATTTCT 300 QY 733 AAGTGTCTGTGATTTGTCACCTTCTTGTCTCTCAACCTTTCAGTTCAGGCCACAGATTC 792 DB 301 AAGTGTCTGTGATTTGTCACCTTCTTGTCTCTCAACCTTTCAGTTCAGGCCACAGATTC 360 QY 793 TATTCAGAGACATGTGGACTTCTGATGCTGCTGAAACACAGCATGCAGTGTATTGGA 852 DB 361 TATTCAGAGACATGTGGACTTCTGATGCTGCTGAAACACAGCATGCAGTGTATTGGA 420 QY 853 AAATATTTTATTAGTATCTTCTTTATAGAGATTTTATTCCCTATTCAATCTGAAT 912

Db 421 AAATATTTTATTAGTATCTTCTTTATAGAGATTTTATTCCCTATTCAATCTGAAT 480 QY 913 TTACACTCAAAATCCACATCACATGTTTGAATCTTTTATTAATCAATACATAA 972 Db 481 TTACAGTCAAAATCCACATCACATGTTTGAATCTTTTATTAATCAATACATAA 540 QY 973 ATCTTCANATGGTATCTTCTAGATGTTTCTTCTCAATGTTTCTGTGGTATCAATTTTC 1032 Db 541 ATCTTCANATGGTATCTTCTAGATGTTTCTTCTCAATGTTTCTGTGGTATCAATTTTC 600 QY 1033 ATGTGGAATTTGTTGATGTTTCCCAATAAACCTTCATAGTGTTCAAAAGCAAAAAA 1092 Db 601 ATGTGGAATTTGTTGATGTTTCCCAATAAACCTTCATAGTGTTCAAAAGCAAAAAA 660 QY 1093 AAAAAAAA 1101 Db 661 AAAAAAAA 669 RESULT 5 AC099183/c LOCUS DEFINITION Rattus norvegicus clone CH230-115J19, *** SEQUENCING IN PROGRESS ***. 66 unordered pieces. AC099183 VERSION AC099183.3 GI:21728919 KEYWORDS HTG; HTGS_PHASE1. SOURCE Norway rat. ORGANISM Rattus norvegicus Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. REFERENCE 1 (bases 1 to 133392) Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayale, M., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonini, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chaviez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, K., Delaney, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Housi, F., Howard, S., Huber, J., Hui, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovat, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichlarge, O., Lieu, C., Liu, J., Liu, W., Louisse, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okuwu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, K., Primus, E., Pu, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savery, G., Scher, S., Scott, G., Shen, H., Shoostrati, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, C., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Unpublished
2 (bases 1 to 133392)
Worley,K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 133392)
Worley,K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17973061.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHAG
Center clone name: CH230-115J19
----- Summary Statistics
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 70190 bases at least Q40
Consensus quality: 73775 bases at least Q30
Consensus quality: 75923 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 66 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1166: contig of 1165 bp in length
* 1265: gap of unknown length
* 2463: contig of 1198 bp in length
* 2464: gap of unknown length
* 2563: gap of unknown length
* 2564: gap of unknown length
* 3747: contig of 1183 bp in length
* 3846: gap of unknown length
* 3847: contig of 1729 bp in length
* 5576: gap of unknown length
* 5676: contig of 1042 bp in length
* 6718: gap of unknown length
* 6818: contig of 1119 bp in length
* 7937: gap of unknown length
* 8036: gap of unknown length
* 9603: contig of 1567 bp in length
* 9604: gap of unknown length
* 9703: gap of unknown length
* 10789: contig of 1086 bp in length
* 10790: gap of unknown length
* 10890: contig of 1008 bp in length
* 11897: gap of unknown length
* 11998: contig of 1387 bp in length
* 13384: gap of unknown length
* 13485: gap of unknown length
* 14726: contig of 1242 bp in length
* 14826: gap of unknown length
* 14827: contig of 1323 bp in length
* 16149: gap of unknown length
* 16249: gap of unknown length
* 16250: contig of 1304 bp in length
* 17553: gap of unknown length
* 17554: contig of 1410 bp in length
* 19063: gap of unknown length
* 19064: contig of 1467 bp in length
* 20630: gap of unknown length
* 20730: contig of 1112 bp in length
* 21842: gap of unknown length
* 21843: contig of 1446 bp in length
*
* 21943: contig of 1056 bp in length
* 22998: gap of unknown length
* 23099: contig of 1064 bp in length
* 24162: gap of unknown length
* 24262: gap of unknown length
* 24263: contig of 1043 bp in length
* 25306: gap of unknown length
* 25405: contig of 1529 bp in length
* 26934: gap of unknown length
* 27034: gap of unknown length
* 27035: contig of 1114 bp in length
* 28148: gap of unknown length
* 28248: gap of unknown length
* 30210: contig of 1962 bp in length
* 30310: gap of unknown length
* 30311: contig of 1945 bp in length
* 32256: gap of unknown length
* 32355: gap of unknown length
* 33390: contig of 1035 bp in length
* 33490: gap of unknown length
* 33491: contig of 1533 bp in length
* 35023: gap of unknown length
* 35123: contig of 1900 bp in length
* 37023: gap of unknown length
* 37123: contig of 1271 bp in length
* 38394: gap of unknown length
* 38494: gap of unknown length
* 38995: contig of 1500 bp in length
* 40094: gap of unknown length
* 41494: contig of 1400 bp in length
* 41594: gap of unknown length
* 42896: contig of 1302 bp in length
* 42996: gap of unknown length
* 45169: contig of 2173 bp in length
* 45170: gap of unknown length
* 46744: contig of 1475 bp in length
* 46745: gap of unknown length
* 46844: gap of unknown length
* 46845: contig of 1664 bp in length
* 48508: gap of unknown length
* 48509: contig of 2177 bp in length
* 50785: gap of unknown length
* 50885: gap of unknown length
* 52950: contig of 2065 bp in length
* 53050: gap of unknown length
* 53511: contig of 2092 bp in length
* 55142: gap of unknown length
* 55242: gap of unknown length
* 55313: contig of 1297 bp in length
* 56339: gap of unknown length
* 56640: contig of 1643 bp in length
* 58282: gap of unknown length
* 58382: gap of unknown length
* 59577: contig of 1195 bp in length
* 59677: gap of unknown length
* 59678: contig of 2050 bp in length
* 61727: gap of unknown length
* 61827: gap of unknown length
* 63172: contig of 1345 bp in length
* 63272: gap of unknown length
* 65192: contig of 1920 bp in length
* 65292: gap of unknown length
* 67342: contig of 2050 bp in length
* 67442: gap of unknown length
* 69603: contig of 2161 bp in length
* 69703: gap of unknown length
* 69704: contig of 2095 bp in length
* 71798: gap of unknown length
* 71898: contig of 2133 bp in length
* 74031: gap of unknown length
* 74131: contig of 1412 bp in length
* 75543: gap of unknown length
* 75643: gap of unknown length
* 77487: contig of 1844 bp in length
* 77587: gap of unknown length
* 79773: contig of 2186 bp in length
* 79873: gap of unknown length
* 82213: contig of 2340 bp in length
* 82313: gap of unknown length
* 85054: contig of 2741 bp in length
* 85154: gap of unknown length
* 88032: contig of 2878 bp in length
* 88132: gap of unknown length
* 89578: contig of 1446 bp in length

```

		* 89579	89678: gap of unknown length	
		* 89679	91651: contig of 1973 bp in length	
Query Match		36.6%;	Score 403.4; DB 2; Length 133392;	
Best Local Similarity		74.8%;	Pred. No. 3.6e-60;	
Matches 539; Conservative		0;	Mismatches 166; Indels 16; Gaps 2;	
QY	6	AACCACCCAGAACCTCGACCCCTCTGAGCGCATATGTCAGCTCTGCTGCTGCTGCTGCTGCTG	65	
Db	66754	AACTCATCCAGAACCTCCA-CCTCTGACACCATGTCAGCTCTGCTGCTGCTGCTGCTGCTG	66696	
QY	66	CTCTGACCAGAGCTGTGTCAGGTCCTCGGCCAGGAGAGCTGCTGCCGCCACAGCTGCTG	125	
Db	66695	CTCTGAGGAGGCTGTGGCCAAG-----GCTGCTGCCAGCCACAGCTGCTG	66651	
QY	126	CCAGACCACTGCTCGAGACCACACTGCTGCGGCCACAGCTGCTGCATTTCCAGTTGCTG	185	
Db	66650	CCAGACCACCTGCTGTAGGACCACCTGCTGCTGCCAGCTGCTGTGTGTCACAGCTGCTG	66591	
QY	186	CAGGCGTTCTGCTGTATCTCCAGTGTCTGAACCCACAGCTGCTGCCTGACACACCTGCTG	245	
Db	66590	CAGACCTAGCTGCTGCCAGCTTTTGTGCTGCGGCCACAGCTGCTGCATTTCTAGCTGCTG	66531	
QY	246	CAGGACCACTGCTGCGGCCACAGTGTGTCATTTCCAGTTGCTGCAGGCGCTTCTGCTG	305	
Db	66530	CAGGCGTCTGCTGCGGCCACAGTTGTTGCATTTCTAGCTGCTGCAGGCGCTTCTGCTG	66471	
QY	306	TATCTCCAGCTGCTGCAAAACCAGCTGCTGTCAGGACCACCTGCTGCCGCCACAGCTGCTG	365	
Db	66470	CCGCCCCACCTGTGCAATTTCTAGCTGCTGTAGGCGTTCTGCTGCCATCCACCTGTTG	66411	
QY	366	CATTTCCAGTGTGCTGAGGCGTTCTGCTGTATCTCCAGTGTCTGCAACCCAGCTGCTG	425	
Db	66410	CATTTCTAGCTGCTGAGGCGTTCTGCTGCGGCCACAGCTGCTGCATTTCTAGTGTGCTG	66351	
QY	426	CAGGACCACTGCTGCGGCCACAGTGTGTCATTTCTAGTTGCTGCAGGCGCTTCTGCTG	485	
Db	66350	TAGGCGTTCTGCTGCGGCCACAGTGTGTCATTTCTAGCTGCTGTAGGCGTTCTGCTG	66291	
QY	486	TATCTCTAGCTGTGTCAAACCAGCTGCTGTCAGACCACTGCTGCCGCCACAGCTGCTG	545	
Db	66290	CCGCCCCAGCTGCTGCATTTCTAGCTGCTGTAGGCGTTCTGCTGCCGTCCACCTGTTG	66231	
QY	546	TATCTCCAGTGTGTACAGGCCCGAGTGTGCCAGCGCTCTGCTGCGCGCCCGGCTTGTG	605	
Db	66230	CATTTCTAGCTGCTGAGGCGTTCTGCTGCGGCCCGACAGCTGCTGCATTTCTAGTGTG	66171	
QY	606	CATTTCTAGTGTGCTGCATTCACAGCTGCTGTGTCCAGCTGCCCGTCCCTTTTCAGTGT	665	
Db	66170	CAGGCGTTCTGTTGCGCGCTTAGCTGTGCTGCAGACCACTGTTGCATTTCTAGTTGCTG	66111	
QY	666	CCGAGCACCTGCTGTAGAACACCTGTCTCCACCCATCTGCTGCGGAGTCTTCTTGTG	725	
Db	66110	CCGCCCCAGCTGCTGCAGACCTCAGTGTGCTGCATCTCCAGCTGCTGCCGCCATCTGTTG	66051	
QY	726	C 726		
Db	66050	C 66050		

RESULT 6

AC098902/c

LOCUS

DEFINITION

2
3
4
5
6
7
8
9
10

ACCESSION
NUMBER

VERSION
KEYWORDS

KEYWORDS
SOURCE

SOURCE ORGANISM

2000

REFERENCE
AUTHORS

1 (bases 1 to 182806)

1 (bases 1 to 182806)
Muzny, D.M., Adams, C., Adio-oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbatia, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buahy, C., Hurch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, K., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, K., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flaeg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Haves, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, H., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Kelly, Y., Johnson, R., Jollivet, S., Joudah, S., Karlsson, E., Kella, Y., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieb, C., Liu, J., Liu, W., Louisguez, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Mashey, E., Mawhinney, E., McLeod, M.P., Meador, M., McI, G., Melzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oraguonye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Peters, L., Peters, L., Pickens, K., Primus, E., Pu, L.L., Quiles, M., Rhee, Y., Rives, M., Rojas, A., Rojoubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabar, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, K., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, Q., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 182806)
Worley, K.C.
Direct Submission
Submitted (05-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 182806)
Worley, K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced q1:17973836.

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1611: contig of 1611 bp in length
* 1612 1711: gap of unknown length
* 1712 2877: contig of 1166 bp in length
* 2878 2977: gap of unknown length
* 2978 4061: contig of 1084 bp in length
* 4062 4161: gap of unknown length
* 4162 5787: contig of 1626 bp in length
* 5788 5887: gap of unknown length
* 5888 7018: contig of 1131 bp in length
* 7019 7118: gap of unknown length
* 7119 8380: contig of 1262 bp in length
* 8381 8480: gap of unknown length
* 8481 10104: contig of 1624 bp in length
* 10105 10204: gap of unknown length
* 10205 11292: contig of 1088 bp in length
* 11293 11392: gap of unknown length
* 11393 12737: contig of 1345 bp in length
* 12738 12837: gap of unknown length
* 12838 14444: contig of 1607 bp in length
* 14445 14544: gap of unknown length
* 14545 15575: contig of 1031 bp in length
* 15576 15675: gap of unknown length
* 15676 17781: contig of 2106 bp in length
* 17782 17881: gap of unknown length
* 17882 19517: contig of 1636 bp in length
* 19518 19617: gap of unknown length
* 19618 21140: contig of 1523 bp in length
* 21141 21240: gap of unknown length
* 21241 22878: contig of 1638 bp in length
* 22879 22978: gap of unknown length
* 22979 24285: contig of 1307 bp in length
* 24286 24385: gap of unknown length
* 24386 26508: contig of 2123 bp in length
* 26509 26608: gap of unknown length
* 26609 28048: contig of 1440 bp in length
* 28049 28148: gap of unknown length
* 28149 30094: contig of 1946 bp in length
* 30095 30194: gap of unknown length
* 30195 31495: contig of 1300 bp in length
* 31496 31594: gap of unknown length
* 31595 34604: contig of 3010 bp in length
* 34605 34704: gap of unknown length
* 34705 36524: contig of 1920 bp in length
* 36525 36724: gap of unknown length
* 36725 38320: contig of 1596 bp in length
* 38321 38420: gap of unknown length
* 38421 39858: contig of 1438 bp in length
* 39859 39859: gap of unknown length
* 39859 43450: contig of 3452 bp in length
* 43451 43550: gap of unknown length
* 43551 46112: contig of 2562 bp in length
* 46113 46212: gap of unknown length
* 46213 48632: contig of 2420 bp in length
* 48633 48732: gap of unknown length
* 48733 50717: contig of 1985 bp in length
* 50718 50817: gap of unknown length
* 50818 53619: contig of 2802 bp in length
* 53620 53719: gap of unknown length
* 53720 56601: contig of 2882 bp in length
* 56602 56701: gap of unknown length
* 56702 59751: contig of 3050 bp in length
* 59752 59851: gap of unknown length
* 59852 62752: contig of 2901 bp in length
* 62753 62852: gap of unknown length

* 62853 66264: contig of 3412 bp in length
* 66265 66364: gap of unknown length
* 66365 70565: contig of 4201 bp in length
* 70566 70665: gap of unknown length
* 70666 74401: contig of 3736 bp in length
* 74402 74501: gap of unknown length
* 74502 79005: contig of 4504 bp in length
* 79006 79105: gap of unknown length
* 79106 84082: contig of 4977 bp in length
* 84083 84182: gap of unknown length
* 84183 87867: contig of 3685 bp in length
* 87868 87967: gap of unknown length
* 87968 93319: contig of 5352 bp in length
* 93320 93420: gap of unknown length
* 93421 97202: contig of 3783 bp in length
* 97203 97302: gap of unknown length
* 97303 101781: contig of 4479 bp in length
* 101782 101881: gap of unknown length
* 101882 107479: contig of 5598 bp in length
* 107480 107579: gap of unknown length
* 107580 111355: contig of 3776 bp in length
* 111356 111455: gap of unknown length
* 111456 115389: contig of 3934 bp in length
* 115390 115489: gap of unknown length
* 115490 119864: contig of 4375 bp in length
* 119865 119964: gap of unknown length
* 119965 124964: contig of 5000 bp in length
* 124965 125064: gap of unknown length
* 125065 129776: contig of 4712 bp in length
* 129777 129876: gap of unknown length
* 129877 138427: contig of 8551 bp in length
* 138428 138527: gap of unknown length
* 138528 145489: contig of 6962 bp in length
* 145490 145589: gap of unknown length
* 145590 154737: contig of 9148 bp in length
* 154738 154837: gap of unknown length
* 154838 164787: contig of 9950 bp in length
* 164788 164887: gap of unknown length
* 164888 173904: contig of 9017 bp in length
* 173905 174004: gap of unknown length
* 174005 182806: contig of 8802 bp in length.

FEATURES

Location/Qualifiers
1..182806

Query Match 36.6%; Score 403.4; DB 2; Length 182806;
Best Local Similarity 74.8%; Pred. No. 3.7e-60;
Matches 539; Conservative 0; Mismatches 166; Indels 16; Gaps 2;
QY 6 AACCCACCTGAGAACTCCACGCTCTGAGCGCCATGCTCAGCTCCTGCTGTGGTCTGCTG 65
DB 57714 AACTCATCCAGAACCTCCA-CCTCTGACACCATGCTCAGCTCCTGCTGTGGTCTGCTG 57656
QY 66 CTCTGACCAAGCTGTGTGCAAGTCTCGGCGAGAGAGCTGCTGCCGCCGAGCTGTG 125
DB 57655 CTCTGAGGAGGCTGTGCGCAAG-----GCTGCTGCCAGCCAGCTGTG 57611
QY 126 CCAGACCACTTCTGCTCAGGACCACTGCTGCCGCCAGCTGCTGCATTTCCAGTCTG 185
DB 57610 CCAGACCACTTCTGCTCAGGACCACTGCTGCCGCCAGCTGCTGCATTTCCAGTCTG 57551
QY 186 CAGGCTTCTCTCTATCTTCTCAGCTGCTGCAACCCAGCTGCTGCCGAGCACTGCTG 245
DB 57550 CAGACCTAGCTGCTGCTCAGCTCTTTGTGCTGCCGCCAGCTGCTGCATTTCCAGTCTG 57491
QY 246 CAGGACCACTTCTGCTGCCGCCAGCTGCTGCAATTTCCAGTCTGCTGCCGCCAGCTG 305
DB 57490 CAGGCTTCTCTGCTGCCGCCAGCTGCTGCAATTTCCAGTCTGCTGCCGCCAGCTG 57431
QY 306 TATCTCCAGCTGCTGCAACCCAGCTGCTGCCGCCAGCTGCTGCCGCCAGCTGCTG 365
DB 57430 CCGGCCCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57371
QY 366 CATTTCACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 425

```
Db 57370 CATTCTAGCTGCTGAGGCTTCTGCTGCGCGCCACCTGTGTGCAATTTCTAGCTGCTG 57311
QY 426 CAGGACCACTGCTGCGCGCCAGCTGCTGATTTCTAGTTGCTGCGAGCCCTTCTGCTG 485
Db 57310 TAGGCTTCTGCTGCGCGCCAGCTGCTGATTTCTAGTGTGCTGAGGCTTCTGCTG 57251
QY 486 TATCTCTAGCTGTTGCAAAACCCAGCTGCTGCGACACCACTGCTGCGCGCCAGCTGCTG 545
Db 57250 CCGCGCCAGCTGCTGATTTCTAGCTGCTGAGGCTTCTGCTGCGCGCCAGCTGCTG 57191
QY 546 TATCTGAGCTGCTACAGGCGCCAGCTGCTGCGACACCACTGCTGCGCGCCAGCTGCTG 605
Db 57190 CATTCTAGCTGCTGAGGCTTCTGCTGCGCGCCAGCTGCTGATTTCTAGCTGCTG 57131
QY 606 CATTCTAGCTGCTGATCCAGCTGCTGCTGCGAGCTGCTGCGCGCCAGCTGCTG 665
Db 57130 CAGGCTTCTGCTGCGCGCCAGCTGCTGCGAGACCACTGCTGATTTCTAGTGTGCTG 57071
QY 666 CCGGACCACTGCTGTAGAACCACTGCTGCGACCACTGCTGCGCGCCAGCTGCTGCTG 725
Db 57070 CCGCGCCAGCTGCTGACACCTGCTGCTGATTTCTAGCTGCTGCGCGCCAGCTGCTG 57011
QY 726 C 726
Db 57010 C 57010
AC067715 192027 bp DNA linear HTG 07-JAN-2001
Mus musculus chromosome 11 clone RP23-326H14, WORKING DRAFT
AC067715 AC067715.24 GI:12025592
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus.
Mus musculus.
Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buhay, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,
Hosak, H., Hou, X., Huber, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loulsegue, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Oguh, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Worley, K. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 192027)
Worley, K.C.
Submitted (27-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 4, 2001 this sequence version replaced gi:1192108.
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: MACE
Center clone name: RP23-326H14
```

```
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 57% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 189676 bases at least Q40
Consensus quality: 191575 bases at least Q30
Consensus quality: 192401 bases at least Q20
Estimated insert size: 189593; sum-of-ctrls estimation.
Quality coverage: 0x in Q20 bases; agarose-fp estimation.
Quality coverage: 7.8x in Q20 bases; sum-of-ctrls estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 6 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 71311: contig of 71311 bp in length
* 71312 71411: gap of unknown length
* 71412 116268: contig of 44857 bp in length
* 116269 116368: gap of unknown length
* 116369 145312: contig of 28944 bp in length
* 145313 145412: gap of unknown length
* 145413 173610: contig of 28198 bp in length
* 173611 173710: gap of unknown length
* 173711 190868: contig of 17158 bp in length
* 190869 190968: gap of unknown length
* 190969 192027: contig of 1059 bp in length.
FEATURES
Location/Qualifiers
1..192027
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-326H14"
BASE COUNT 55026 a 41185 c 39554 g 55749 t 513 others
ORIGIN
Query Match 34.4%; Score 378.2; DB 2: Length 192027;
Best Local Similarity 76.7%; Pred. No. 8.7e-56;
Matches 475; Conservative 0; Mismatches 143; Indels 1; Gaps 1;
QY 2 TGGGAACCTACCCAGAACCTCCACCTCTGAGCCAGCTGCTGCTGCTGCTGCTGCTG 61
Db 33817 TCTAAACCCATCCAGAACCTCCA-CCTCTGACACCATGGCCAACTCTTCTGCTGCTG 33875
QY 62 TCTGCTCTGACACAGAGCTGTGCTCAAGCTCTCGGCGCAGAGAGCTGCTGCCGAGCT 121
Db 33876 TCTGCTCTGAGAGAGAGCTGTGCGCAAGCTGCTGCGCAGCCAGCTGCTGCCAGAGCACT 33935
QY 122 GCTGCGCAGACCACTGCTGCGCAGCACTGCTGCGCGCCAGCTGCTGCTGCTGCTGCT 181
Db 33936 GCTGTAGGACCACTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33995
QY 182 GCTGAGGCTTCTGCTGCTGCTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCCGAGCACT 241
Db 33996 GCTGCCAGTCTGTGCTGCTGCCAGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34055
QY 242 GCTGAGGACCACTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301
Db 34056 GCTGAGGCGCCAGCTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34115
QY 302 GCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCGAGGACCACTGCTGCCGCGCACT 361
Db 34116 GCTGTGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34175
QY 362 GCTGTATTTTCCAGTGTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
Db 34176 GCTGTATTTTCCAGTGTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 34235
```


QY 422 GCTGCAGGACACCTGCTGCCGCCAGCTGCTGCATTTCTAGTTGCTGCAGGCTTCT 481
 Db 34236 GCTGCAGGCTTCTGCTGCCGCCAGCTGCTGCAGACCTAGCTGTTGCATTTCTAGCT 34295
 QY 482 GCTGTATCTCTAGCTTGTTCACAAACAGCTGCTGTCAGACACCTGCTGCCGCCAGCT 541
 Db 34296 GCTGCCGCCAGCTGCTGTTGTCAGCTGCTGCAGACCCAGCTGCTCATCTCCAGCT 34355
 QY 542 GCTGTATCTCTAGCTGCTACAGGCCAGCTGCTGTCAGACCCCTGCTGCCGCCAGCT 601
 Db 34356 GCTGCCGCCAGCTGCTGTCAGACACCTGCTGTCAGGACACCTGCTACCGCCAGCT 34415
 QY 602 GCTGCATTTCTAGTTGCTG 620
 Db 34416 GCTCTAGTGGTCTTCTGCTG 34434

RESULT 8
 AL590997/c
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP23-132J20 on chromosome 11, complete sequence.
 ACCESSION AL590997
 VERSION AL590997.13 GI:17221205
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 200602)
 Blakey,S.
 Direct Submission
 Submitted (29-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Nov 30, 2001 this sequence version replaced gi:16973923.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-132J20 is from the RPI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6

FEATURES
 source
 Location/Qualifiers
 1..200602
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-132J20"
 /clone_lib="RPI-23"
 42963
 /note="Tandem repeat. Forced join. Cap size estimated to be approximately 260bp by restriction digest data."
 115069..115073
 misc_feature
 misc_feature

/note="1327 bases of IS2 transposon (V0061) removed here. Flanking repeat sequences are irregular."
 115074..115111
 /note="Sequence from overlapping clone RP23-212C14 (AL590992). Assembly confirmed by restriction digest."
 188825..188863
 /note="Sequence from uni-directional dGTP big dye terminator reads only."
 BASE COUNT 58430 a 40464 c 42904 g 58804 t
 ORIGIN
 Query Match 34.4%; Score 378.2; DB 10; Length 200602;
 Best Local Similarity 76.7%; Pred. No. 8.8e-56;
 Matches 475; Conservative 0; Mismatches 143; Indels 1; Gaps 1;
 QY 2 TGGGAATCGACCCAGAACCTCCACCTCTGAGCCAGGAGCTGCTGCCGCCAGCT 121
 Db 55147 TGTAAACCCATCCAGAACCTCCA-CCTCTGACACCATGCCAACTCTTGTGCTCTG 55089
 QY 62 TCTGCTCTGACACAGAGCTGCTCAAGGTCTCGGCCAGGAGAGCTGCTGCCGCCAGCT 121
 Db 55088 TCTGCTCTGAGAGAGCTGTGGCCAGGCTGTGCCAGCCAGCTGCTGCCAGCACT 55029
 QY 122 GCTGCAGACACCACTGCTGCAGGAGCACCTGCTGCCGCCAGCTGCTGCCAGCT 181
 Db 55028 GCTGTAGGACCACTGCTGCCGCCAGCTGCTGTGTCTCCAGCTGCTGCAGACCT 54969
 QY 182 GCTCAGGCTTCTGCTGCTATCTCCAGCTGTGCAAAACCACTGCTGCTGAGCACT 241
 Db 54968 GCTGCCAGTCTGCTGCTGCCAGCCACCTGCTGCCGCCAGCTGCTGCATTTCTAGCT 54909
 QY 242 GCTCAGGACCACTGCTGCCGCCAGCTGCTGCCAGCTGCTGCCAGCTGCTGCCAGCT 301
 Db 54908 GCTCAGGCCCACTGCTGCCGCCAGCTGCTGCTATTTCCAGCTGCTGCAGGCT 54849
 QY 302 GCTGTATCTCCAGCTGCTGCAAAACCACTGCTGCAGGACCACTGCTGCCGCCAGCT 361
 Db 54848 GCTGTGCCGCCAGCTGCTGCATTTCCAGCTGCTGCAGGCCCACTGCTGCCGCCAGCT 54789
 QY 362 GCTGCATTTCTGCTGCTGAGGCTTCTGCTGATCTCCAGCTGCTGCTGCAAAACCACT 421
 Db 54788 GCTGTATTTCTAGCTGCTGCAGGCTTCTGCTGCCGCCAGCTGCTGCATTTCTAGCT 54729
 QY 422 GCTCAGGACCACTGCTGCCGCCAGCTGCTGCCAGCTGCTGCTGCCAGCTTCTAGCT 481
 Db 54728 GCTCAGGCTTCTGCTGCCGCCAGCTGCTGCAGACCTAGCTGCTGCTGCTAGCT 54669
 QY 482 GCTGTATCTCTAGCTGCTTGCACCAACCACTGCTGCCAGACCACTGCTGCCGCCAGCT 541
 Db 54668 GCTGCCGCCAGCTGCTGCTGCTCCAGCTGCTGCAGACCCCACTGCTGCATTTCTAGCT 54609
 QY 542 GCTGTATCTCCAGCTGCTACAGGCCCACTGCTGCCAGGCCCTGCTGCTGCCGCCAGCT 601
 Db 54608 GCTGCCGCCCACTCTCTTCCAGACCACTGCTGCAGGACCACTGCTGCTGCCGCCAGCT 54549
 QY 602 GCTGCATTTCTAGTCTGCTG 620
 Db 54548 GCTCTAGTGGTCTTCTGCTG 54530
 RESULT 9
 AL646094
 LOCUS
 DEFINITION Mus musculus chromosome 11 clone RP23-142E7, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AL646094
 VERSION AL646094.6 GI:19031714
 KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1

AUTHORS	Ellington,A.
TITLE	Direct Submission
JOURNAL	Submitted (26-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Feb 28, 2002 this sequence version replaced gi:18135257. ----- Genome Center ----- Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information ----- Center project name: bm142E7 ----- Summary Statistics ----- Assembly program: XGAP4; version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 99% of reads Chemistry: Dye-primer Big Dye; 0% of reads Consensus quality: 192882 bases at least Q40 Consensus quality: 192997 bases at least Q30 Consensus quality: 193039 bases at least Q20 Insert size: 197658; sum-of-contigs Insert size: 203092; 2.3% error; agarose-fp Quality coverage: 13.20x in Q20 bases; sum-of-contigs Quality coverage: 13.17x in Q20 bases; agarose-fp -----
COMMENT	* NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.
FEATURES	Location/Qualifiers 1..197658 /organism="Mus musculus" /db_xref="taxon:10090" /chromosome="11" /clone="RP23-142E7" /clone_lib="RPCI-23" misc_feature 1..197658 /note="assembly_fragment:02832"
BASE COUNT	59207 a 42073 c 41216 g 55162 t
ORIGIN	
Query Match	33.4%; Score 368.2; DB 2; Length 197658;
Best Local Similarity	72.7%; Pred. No. 4.8e-54;
Matches 524; Conservative	0; Mismatches 178; Indels 19; Gaps 3;
QY	6 AACCACCCAGAACCTCCACCTCTGCAGCATGGTCAGCTCCTCGTGTGCTGCTGCTG 65
Db	17709 AACCTACCAAGAATCTCACCTTTAAACCAATGTCAGCTCCCTGTTGTGCTGTGCTG 17767
QY	66 CTCTCAGCAGAGCTGTGTGAAGTCTCGGCCAGAGAGCTGCTGCCGCCCCAGCTGCTG 125
Db	17768 CTTGAGGAGGCTGTGCCAAG-----GCTGCTGCCAGCCAGCTGCTG 17812
QY	126 CCAGACCACTGCTGCAGGACACCTGCTGCCGCCCCAGCTGCTGCATTCCAGTTCCTG 185
Db	17813 CCAGACTACCTGCTTAGGACACCTGCTGCCGCCCCAGCTGCTGTGTGCCAGCTGCTG 17872
QY	186 CAGGCTTCCTGCTGTATCTCCAGCTGCTGCAACCCAGCTGCTGCCTGACCCAGCTGCTG 245
Db	17873 CAGGCCAGCTGCTGTGTGTCCAGCTGCTGCAGACCCAGTGTTGCCAGTGTGTGCTGCTG 17932
QY	246 CAGGACCACTGCTGCCGCCCCAGCTGCTGCATTTCAGTTGCTGCAGGCTCTTCTGCTG 305
Db	17933 CCAGCCCAGCTGCTGCCGCCCCAGCTGCTGTGCCGCCAGCTGCTGCATTTCAGTTCCTG 17992
QY	306 TATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGACCACTGCTGCCGCCCCAGCTGCTG 365
Db	17993 TAGACCTTCTGCTCCGC---CCAGCTGCTGTGTGTGCCAGCTGCTGCAGGCCCCAGCTGCTG 18049
QY	366 CATTTCCAGTTCTGCAGCCTTCTGCTGTATCTTCAGCTGCTGTCGAACCCAGCTGCTGCT 425
Db	18050 CCAGTCTGTGTGCTGCAGCCACCTGCTGCCGCCCCAGCTGCTGTGCTGCCCCAGCTGCTG 18109

```
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-132J20"
/clone_11b="RPC1-23"
42963
/note="Random repeat. Forced join. Gap size estimated to
be approximately 260bp by restriction digest data."
misc_feature 115069..115073
/note="1327 bases of IS2 transposon (V0061) removed here.
Flanking repeat sequences are irregular."
misc_feature 115074..115111
/note="Sequence from overlapping clone RP23-212C14
(AL590992). Assembly confirmed by restriction digest."
misc_feature 189825..189863
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
BASE COUNT 58430 a 40464 c 42904 g 58804 t
ORIGIN

Query Match 33.4%; Score 368.2; DB 10; Length 200602;
Best Local Similarity 72.7%; Pred. No. 4.8e-54;
Matches 524; Conservative 0; Mismatches 178; Indels 19; Gaps 3;

QY 6 AACCCACCAGAACCTCCACCTCTGACGCCATGCTCAGCTCCTGCTGCTGCTGCTG 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174439 AACCTACCAAGACCTTCACC-TCTAAACCATGCTCAGCTCCTGTTGTGCTGCTG 174497

QY 66 CTCTGACAGAGCTGTGTTGAAGTCTCGGCCAGGAGAGCTGTCGCCGCCACGCTGCTG 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174498 CTCTGAGAGAGGCTGTGGCCAAAG-----GCTGCTGCCAGCCACGCTGCTG 174542

QY 126 CCAGACCACTGCTGCGAGGACACCTGCTGCCGCCAGCTGCTGCTGCTGCTGCTG 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174543 CCAGACTACCTGCTGTAGGACACACTGCTGCCGCCAGCTGCTGCTGCTGCTGCTG 174602

QY 186 CAGGCTTCTCTGCTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACCACTGCTG 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174603 CAGGCCACAGCTGCTGTGTCCAGCTGCTGCGAGACCCAGTGTGCCAGTCTGTGCTG 174662

QY 246 CAGGACCACTGCTGCCGCCACAGCTGCTGATTTCCAGCTGCTGCGAGCCCTTCTGCTG 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174603 CCAGCCCACTGCTGCCGCCACAGCTGCTGCGCCAGCTGCTGCAATTTCTAGCTGCTG 174722

QY 306 TATCTCCAGCTGCTGCAAAACCCAGCTGCTGCGAGACCACTGCTGCCGCCACGCTGCTG 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174723 TAGACCTTGTGCTGCGC-----CCAGCTGCTGTGTGCCAGCTGCTGCGAGCCCACTGCTG 174779

QY 366 CATTTCCAGTTGCTGCGAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTG 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174780 CCAGTCTGTGTGCTGCCACCACTGCTGCTGCCGCCAGCTGCTGCCGCCACCTGTTG 174839

QY 426 CAGGACCACTGCTGCGGCCACAGCTGCTGATTTCTAGTTGCTGCGAGGCTTCTGCTGCTG 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174840 CATTTCCAGCTGCTGCGGCCCTGCTGTGTGATTTCCAGCTGCTGTGATATAGCTGCTG 174899

QY 486 TATCTCTAGCTGTGCAAAACCCAGCTGCTGCTGCGAGACCACTGCTGCCGCCACGCTGCTG 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174900 CGGCCCCACCTGTTGCAATTTCCAGCTGCTGCGGCCCACTGCTGCGAGGCCAGCTGCTG 174959

QY 546 TATCTCCAGCTGCTACAGGCCCTGCTGCTGCGAGCCCTGCTGCTGCCGCCCGGCTTGTGCTG 605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174960 TGTGTCCAGCTGCTGCGAGACCCAGTGTGATCTCCAGCTGCTGCGGCCCTTCTGCTGCTG 175019

QY 606 CATTTCTAGTTGCTGCTCATCCAGCTGCTGTGTGTTCCAGCTGCGCGCTGCCCTTTTCAGCTG 665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175020 TAGTTCTAGCTGCTGTGATCTAGCTGCTGCCGCCACGCTGCTGCGAGGCCAGCTGCTG 175079

QY 666 CCCGACCACTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCCGCCAGTCTTCTGCTG 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175080 TGTGTCCAGCTGCTGCGAGACCCAGTGTGCTGATCTTAGCTGCTGCCGCCACCTGTTG 175139

QY 726 C 726
    |
```

Db 175140 C 175140

RESULT 11

AC025904

LOCUS

DEFINITION

AC025904 Homo sapiens clone RP11-585G19, WORKING DRAFT SEQUENCE, 17 unordered pieces.

AC025904 Homo sapiens.

AC025904.2 GI:7382590

HTG; HTGS_PHASE1; HTGS_DRAFT.

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 174032)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens, clone RP11-585G19

Unpublished

2 (bases 1 to 174032)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,I., Boukhgaller,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,C., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karakas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPherson,K., Melcham,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triplino,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,P., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 1, 2000 this sequence version replaced gi:7249428.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8441

Center clone name: 585.G.19

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 158771 bases at least Q40

Consensus quality: 167389 bases at least Q30

Consensus quality: 170660 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 172432; sum-of-contigs

Quality coverage: 3.9 in Q20 bases; agarose-fp

Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2588: contig of 2588 bp in length
* 2589 2688: gap of 100 bp
* 2689 6465: contig of 3777 bp in length
* 6466 6565: gap of 100 bp
* 6566 8846: contig of 2281 bp in length
* 8847 8946: gap of 100 bp
* 8947 13104: contig of 4158 bp in length
* 13105 13204: gap of 100 bp
* 13205 19430: contig of 6226 bp in length
* 19431 19530: gap of 100 bp
* 19531 25564: contig of 6034 bp in length
* 25565 25664: gap of 100 bp
* 25665 31377: contig of 5713 bp in length
* 31378 31477: gap of 100 bp
* 31478 39523: contig of 8046 bp in length
* 39524 39623: gap of 100 bp
* 39624 47042: contig of 7419 bp in length
* 47043 47142: gap of 100 bp
* 47143 55411: contig of 8269 bp in length
* 55412 55511: gap of 100 bp
* 55512 66000: contig of 10489 bp in length
* 66001 66100: gap of 100 bp
* 66101 74184: contig of 8084 bp in length
* 74185 74284: gap of 100 bp
* 74285 86712: contig of 12428 bp in length
* 86713 86812: gap of 100 bp
* 86813 103527: contig of 16715 bp in length
* 103528 103627: gap of 100 bp
* 103628 120465: contig of 16838 bp in length
* 120466 120565: gap of 100 bp
* 120566 145510: contig of 24945 bp in length
* 145511 145610: gap of 100 bp
* 145611 174032: contig of 28422 bp in length.

FEATURES

Location/Qualifiers
1..174032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-585G19"
/clone_lib="RP11-11 Human Male BAC"
1..2588
/note="assembly_fragment"
vector_side:right
clone_end:T7
2689..6465
/note="assembly_fragment"
6566..8846
/note="assembly_fragment"
8947..13104
/note="assembly_fragment"
13205..19430
/note="assembly_fragment"
19531..25564
/note="assembly_fragment"
25665..31377
/note="assembly_fragment"
31478..39523
/note="assembly_fragment"
39624..47042
/note="assembly_fragment"
47143..55411
/note="assembly_fragment"
55512..66000
/note="assembly_fragment"
66101..74184
/note="assembly_fragment"
74285..86712
/note="assembly_fragment"
86813..103527
/note="assembly_fragment"
103628..120465

misc_feature /note="assembly_fragment"
120566..145510
misc_feature /note="assembly_fragment"
145611..174032
misc_feature /note="assembly_fragment"
clone_end:SP6
vector_side:right
BASE COUNT 50722 a 34581 c 34553 g 52576 t 1600 others
ORIGIN
Query Match 32.3%; Score 356; DB 2; Length 174032;
Best Local Similarity 69.4%; Pred. No. 6.2e-52;
Matches 504; Conservative 0; Mismatches 210; Indels 12; Gaps 1;
QY 1 CTGGGAACCCAGCCAGAAACCTCTACACCTCTGACAGCCCATGGTCAGCTCCTGCTGTGATCT 60
DB 6672 CTGGAACCCAGCCAGAAACCTCTACACCTCTGACAGCCCATGGTCAGCTCCTGCTGTGATCT 6731
QY 61 GTCAGCTCTGACAGAGAGCTGTGGTCAAGGTCCTGGCCAGGAGAGCTGCTGCGGCGGAGC 120
DB 6732 GTGTGCTCTGACAGAGAGCTGT-----GGCCTAGAGAACTGCTGTGAGAGAGCAGC 6779
QY 121 TGCTGCCAGAGCCAGCTGTGTCAGAGACCACTGCTGCGGCCGCCAGCTGCTGTGATTTGAGT 180
DB 6780 TGCTGCCAGAGCCAGCTGTGTCAGAGACCACTGCTGCGGCCGCCAGCTGCTGTGATTTGAGT 6839
QY 181 TGCTGCCAGAGCTTCTGCTGTATCTCCAGCTGTGTCAGAAACCCAGCTGCTGTGATTTGAGT 240
DB 6840 TGCTGCCAGAGCTTCTGCTGTATCTCCAGCTGTGTCAGAAACCCAGCTGCTGTGATTTGAGT 6899
QY 241 TGCTGCCAGAGCCAGCTGTGCGGCCGCCAGCTGCTGCGGCCGCCAGCTGCTGTGATTTGAGT 300
DB 6900 TGCTGTGTGTCAGCTGTGTCAGAGCCCACTGCTGCGGCCGCCAGCTGCTGTGATTTGAGT 6959
QY 301 TGCTGTATCTCCAGCTGTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 360
DB 6960 TGCTGCCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 7019
QY 361 TGCTGTATCTCCAGCTGTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 420
DB 7020 TGCTGCCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 7079
QY 421 TGCTGCCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 480
DB 7080 TACTGTGTGTCAGCTGTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 7139
QY 481 TGCTGTATCTCCAGCTGTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 540
DB 7140 TGCTGCCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 7199
QY 541 TGCTGTATCTCCAGCTGTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 600
DB 7200 TGCTGTATCTCCAGCTGTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 7259
QY 601 TGCTGTATCTCCAGCTGTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 660
DB 7260 TATTGCAATTTGAGCTGTGTCAGAGCCCACTGCTGTCAGAGCCCACTGCTGTGATTTGAGT 7319
QY 661 AGCTGCCAGAGCCCACTGCTGTGAGAACCACTGCTTCCACCCCACTGCTGTGATTTGAGT 720
DB 7320 TGCCTGTGCTTTCAGCTGCTGTCAGAGCCCACTGCTTCCCTGTCATCTCCAGTGTGTCGCGGCGGAGC 7379
QY 721 TGCTGC 726
DB 7380 TGCTGC 7385
RESULT 12
AC107303/c AC107303 347572 bp DNA linear HTG 09-MAY-2002
LOCUS Homo sapiens chromosome 3 clone RP11-595122, WORKING DRAFT
DEFINITION SEQUENCE, 32 unordered pieces.
ACCESSION AC107303

RESULT 13																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
-----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

```
Db 233 TGCTGTGTCTCCAGCTGCTGCAGACCCAGTGTGTCAGTCTGTGTGCTGCCAGCCACC 292
Qy 301 TGCTGTATCTCCAGCTGCTGCAAAACCAGTGTGTGCAGACCCAGTGTGCTGCCGCCAGC 360
Db 293 TGCTGCAGACCCCAATGCTGCCAGACTACCTGTGTGTAGGACCACTGCTGCCGCCAGC 352
Qy 361 TGCTGCATTTTCCAGTGTCTGCAGGCCCTTCTGCTGTGTATCTCCAGTGTCTGCAAAACCAGC 420
Db 353 TGCTGCAGGCCCCAGTGTCTGCCAGTGTGTGTGTGCCAGCCCACTGCTGCTGCCCCAGC 412
Qy 421 TGCTGCAGGACCTGCTGCCGCCCCAGTGTGTGTGTGTGTATTTCTAGTTGCTGCAGGCCCTTC 480
Db 413 TACTGTGTGTCTCCAGTGTCTGCAGACCCAGTGTGTGCCAGACCACTGCTGCAGAAACCACC 472
Qy 481 TGCTGTATCTAGTGTGTGCAAAACCAGTGTGTGCCAGACCACTGCTGCCGCCAGC 540
Db 473 TGCTGCCGCCCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 532
Qy 541 TGCTGTATCTCCAGTGTCTACAGGCCCCAGTGTGTGCCAGGCCCTGCTGCTGCCGCCGGCT 600
Db 533 TGCTGTAGCCCTCTGCTGGCAAAACCACCTGCTGTACAGGACCACTGCTGCCACCCAGC 592
Qy 601 TGCTGCATTTTCTAGTGTGTATCTCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
Db 593 TATTGCATTTTCCAGTGTCTGCCAGCCCTTCTGTGCAGTATCTCTAGCAGCAGTAGTCTCTCC 652
Qy 661 AGCTGCCGCCAGCCTGCTGTGTAGAACACCTGCTTCCACCCCATCTGCTGCCGCCAGTTCT 720
Db 653 TGCTGTGGTCTACAGTGTCTGCAGGCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 712
Qy 721 TGCTGC 726
Db 713 TGCTGC 718

RESULT 14
LOCUS HSA406939 1203 bp mRNA linear PRI 31-JAN-2001
DEFINITION Homo sapiens mRNA for keratin associated protein 4.7 (KRTAP4.7
gene).
ACCESSION AJ406939
VERSION AJ406939.1 GI:12655451
KEYWORDS keratin associated protein 4.7; KRTAP4.7 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1203)
Rogers, M.A., Langbein, L., Winter, H., Elmann, C., Korn, B. and
Schweizer, J.
Characterization of a cluster of human high/ ultrahigh keratin
associated proteins on chromosome 17q12-21
Unpublished
JOURNAL REFERENCE 2 (bases 1 to 1203)
AUTHORS Rogers, M.A.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2000) Rogers M.A., Research Program B, German
Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg,
Germany 69120, GERMANY
COMMENT Related genomic sequence: AC025904 (142457-143089nt, neg strand).
FEATURES
Location/Qualifiers
1..1203
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q12-21"
/clone="sc1181"
/tissue_type="scalp"
47..679
/gene="KRTAP4.7"
47..679
/gene="KRTAP4.7"
/function="structural protein"

polyA_signal 242 a 389 c 248 g 324 t
BASE COUNT 1127. 1132
/standard_name="KAP4.7"
/codon_start=1
/product="keratin associated protein 4.7"
/protein_id="CAC27578.1"
/db_xref="GI:12655452"
/translation="MVSSCCGSCVSDGGQGVLCQETCCRPSCCTTCCCTTCYRPSC
CVSSCCRPCCQSVCCPTCCCTTCHPRCCITSSCCRPSCCMSCCKKPCQCS
VCCOPTCCHPSCCISCCRPSCCVSRCCRPCCQSVCCOPTCCRPSCISCCRPSCC
ESSCCRPCCRPCCCLRPVCGRVSCHTTCYRPTCVISTCPRPICCASSCC"

Query Match 32.2% Score 354.4 DB 9; Length 1203;
Best Local Similarity 71.0% Pred. No. 7.9e-52;
Matches 491; Conservative 0; Mismatches 186; Indels 15; Gaps 1;

Qy 2 TGGAAACCCACCCAGAACCTCCACCCCTCTTGACGCCATGTCAGTCTCTGCTGTGATCTG 61
Db 12 TGAACACCCACCCAGATCTCTCCCGTTCGACACCATGGTCAGTCTCTGCTGTGATCTG 71
Qy 62 TCTGCTCTGACACAGAGTGTGTCAAGTCTGCGCCAGGAGAGTGTGCTGCCGCTGCTGAT 121
Db 72 TGTGCTCTGACACAGAGTGTGTCAAGTCTGCGCCAGGAGAGTGTGCTGCCGCTGCTGAT 131
Qy 122 GCTGCCAGACACCTGTGTGACGACACCTGCTGCCGCCAGTGTGCTGTGATCTGCTGAT 181
Db 132 GCTGTACAGACACCTGTGTGACGACACCTGCTGCCGCCAGTGTGTGCTGTGATCTGCT 191
Qy 182 GCTGCCAGGCTTCTGCTGTATCTTCCAGTGTGCTGCAAAACCAGTGTGCTGTGATCTG 241
Db 192 GCTGCCAGGCTTCTGCTGTATCTTCCAGTGTGCTGCAAAACCAGTGTGCTGTGATCTG 251
Qy 242 GCTGCCAGGACACCTGCTGTGCGGCCAGTGTGCTGATCTTCCAGTGTGCTGTGATCTGCT 301
Db 252 GCTGTGAGAGACCTGCTGTGCAACCCAGTGTGCTGCAAAACCAGTGTGCTGTGATCTGCT 311
Qy 302 GCTGTATCTCCAGTGTGCAAAACCAGTGTGCTGCAAAACCAGTGTGCTGTGATCTGCT 361
Db 312 GCTGTATCTCCAGTGTGCAAAACCAGTGTGCTGCAAAACCAGTGTGCTGTGATCTGCT 371
Qy 362 GCTGCATTTCCAGTGTGCTGTGAGGCTTCTCTGCTGTATCTCCAGTGTGCTGTGATCTG 421
Db 372 GCTGCCACCCAG-----CTGCTGCATCTCCAGTGTGCTGTGATCTGCTGTGATCTG 416
Qy 422 GCTGCCAGACACCTGCTGTGCGGCCAGTGTGCTGATCTTCCAGTGTGCTGTGATCTGCT 481
Db 417 GCTGTGTGCCAGTGTGCTGAGGCCAGTGTGCTGCCAGTGTGCTGTGATCTGCTGTGATCT 476
Qy 482 GCTGTATCTCTAGTGTGTGCAAAACCAGTGTGCTGCCAGACCACTGCTGCCGCCAGTGT 541
Db 477 GCTGCCGTGCCAGTGTGCTGATCTTCCAGTGTGCTGCCGCCAGTGTGCTGTGATCTGCT 536
Qy 542 GCTGTATCTCCAGTGTGTACAGGCCAGTGTGCTGCCAGCCCTCTGCTGTGATCTGCTGTG 601
Db 537 GCTGCCGCCATGCTGTGCGGCCAGTGTGCTGCCAGTGTGCTGTGATCTGCTGTGATCTG 596
Qy 602 GCTGCATTTCTAGTGTGTGTCATCCAGTGTGCTGTGTCAGTGTGCTGTGATCTGCTGTG 661
Db 597 CTTGCCACACCACTGTGATCTGCGCCAACTGCTGTGATCTTCCAGTGTGCTGTGATCTGCT 656
Qy 662 GCTGCCGCCAGCCTGCTGTGTAGAACACCACTGCTGCCGCCAGTGTGCCGCCAGTGT 693
Db 657 GCTGTGCTCTCTTGTGCTGTGCTGAGTGCACCTGCTGCCGCCAGTGTGCCGCCAGTGT 688

RESULT 15
AC037482/2
LOCUS AC037482 186078 bp DNA linear PRI 23-FEB-2002
DEFINITION Homo sapiens chromosome 17, clone RP11-619M15, complete sequence.
KEYWORDS AC037482
ACCESSION AC037482
VERSION AC037482.14 GI:18873872
KEYWORDS HTG.
```

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (23-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 23, 2002 this sequence version replaced gi:18693527.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRK
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9334
Center clone name: 619_M15

Only the first 186.1 kilobases of this clone are being submitted.
The remainder overlaps accession number AC004231 [WICGR project L297].

FEATURES	source	location/Qualifiers
repeat_region	1. .186078	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="17" /map="17"
repeat_region	complement(1. .313)	/clone="RP11-619M15"
repeat_region	complement(314. .642)	/rpt_family="L1"
repeat_region	complement(1000. .1127)	/rpt_family="L1MC5"
repeat_region	1154. .1465	/rpt_family="L1MC4a"
repeat_region	complement(1504. .1585)	/rpt_family="AluSg1"
repeat_region	complement(1707. .2219)	/rpt_family="L1MA9"
repeat_region	complement(2278. .2566)	/rpt_family="L1MC4a"
repeat_region	complement(2682. .2944)	/rpt_family="AluX"
repeat_region	2953. .2983	/rpt_family="L1PA8"
repeat_region	3070. .3231	/rpt_family="A1_rich"
repeat_region	4320. .4877	/rpt_family="MIR"
repeat_region	complement(5609. .6042)	/rpt_family="L1M2b"
repeat_region	6499. .6580	/rpt_family="L1PA16"
unsure	complement(6526. .6560)	/rpt_family="(TTCCG)n"
repeat_region	6581. .6704	/note="single clone coverage"
repeat_region	complement(7231. .7419)	/rpt_family="LTR16a"
repeat_region	7597. .7776	/rpt_family="L1PA5"
repeat_region	complement(8190. .8513)	/rpt_family="LTR16c"
repeat_region	complement(8190. .8513)	/rpt_family="L12"

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRK
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9334
Center clone name: 619_M15

Only the first 186.1 kilobases of this clone are being submitted.
The remainder overlaps accession number AC004231 [WICGR project L297].

FEATURES	source	location/Qualifiers
repeat_region	1. .186078	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="17" /map="17"
repeat_region	complement(1. .313)	/clone="RP11-619M15"
repeat_region	complement(314. .642)	/rpt_family="L1"
repeat_region	complement(1000. .1127)	/rpt_family="L1MC5"
repeat_region	1154. .1465	/rpt_family="L1MC4a"
repeat_region	complement(1504. .1585)	/rpt_family="AluSg1"
repeat_region	complement(1707. .2219)	/rpt_family="L1MA9"
repeat_region	complement(2278. .2566)	/rpt_family="L1MC4a"
repeat_region	complement(2682. .2944)	/rpt_family="AluX"
repeat_region	2953. .2983	/rpt_family="L1PA8"
repeat_region	3070. .3231	/rpt_family="A1_rich"
repeat_region	4320. .4877	/rpt_family="MIR"
repeat_region	complement(5609. .6042)	/rpt_family="L1M2b"
repeat_region	6499. .6580	/rpt_family="L1PA16"
unsure	complement(6526. .6560)	/rpt_family="(TTCCG)n"
repeat_region	6581. .6704	/note="single clone coverage"
repeat_region	complement(7231. .7419)	/rpt_family="LTR16a"
repeat_region	7597. .7776	/rpt_family="L1PA5"
repeat_region	complement(8190. .8513)	/rpt_family="LTR16c"
repeat_region	complement(8190. .8513)	/rpt_family="L12"

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRK
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9334
Center clone name: 619_M15

Only the first 186.1 kilobases of this clone are being submitted.
The remainder overlaps accession number AC004231 [WICGR project L297].

FEATURES	source	location/Qualifiers
repeat_region	1. .186078	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="17" /map="17"
repeat_region	complement(1. .313)	/clone="RP11-619M15"
repeat_region	complement(314. .642)	/rpt_family="L1"
repeat_region	complement(1000. .1127)	/rpt_family="L1MC5"
repeat_region	1154. .1465	/rpt_family="L1MC4a"
repeat_region	complement(1504. .1585)	/rpt_family="AluSg1"
repeat_region	complement(1707. .2219)	/rpt_family="L1MA9"
repeat_region	complement(2278. .2566)	/rpt_family="L1MC4a"
repeat_region	complement(2682. .2944)	/rpt_family="AluX"
repeat_region	2953. .2983	/rpt_family="L1PA8"
repeat_region	3070. .3231	/rpt_family="A1_rich"
repeat_region	4320. .4877	/rpt_family="MIR"
repeat_region	complement(5609. .6042)	/rpt_family="L1M2b"
repeat_region	6499. .6580	/rpt_family="L1PA16"
unsure	complement(6526. .6560)	/rpt_family="(TTCCG)n"
repeat_region	6581. .6704	/note="single clone coverage"
repeat_region	complement(7231. .7419)	/rpt_family="LTR16a"
repeat_region	7597. .7776	/rpt_family="L1PA5"
repeat_region	complement(8190. .8513)	/rpt_family="LTR16c"
repeat_region	complement(8190. .8513)	/rpt_family="L12"

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRK
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9334
Center clone name: 619_M

Result No.	Score	Query			DB	ID	Description
		Match	Length	ε			
1	701.5	47.3	186	2	A45910	ultra-high-sulfur	
2	693.5	46.7	223	2	B38346	ultra-high-sulfur	
3	684	46.1	230	2	A38346	ultra-high-sulfur	
4	619	41.7	151	2	S60314	hair keratin cysteine-rich hair	
5	569.5	38.4	126	2	I46489	high-sulfur protein	
6	560	37.7	188	2	JC6547	high-sulfur keratin	
7	463.5	31.2	175	2	S37649	high-sulfur keratin	
8	451	30.4	177	2	S37650	keratin high-sulfur	
9	434	29.2	182	1	KRSHHD	high-sulfur wool m	
10	432	29.1	182	2	I47105	ultra high-sulfur	
11	431.5	29.1	169	1	S18946	keratin high-sulfu	
12	408.5	27.5	172	1	KRSHHA	high-sulfur wool m	
13	402	27.1	162	2	I47106	high-sulfur wool m	
14	384.5	25.9	162	2	I47107	high-sulfur wool m	
15	375	25.3	654	2	T30136	hypothetical prote	
16	360.5	24.3	152	2	I47108	high-sulfur wool m	
17	358.5	24.2	152	1	KRSHHC	keratin high-sulfu	
18	358.5	24.2	152	2	I47111	high-sulfur wool m	
19	358	24.1	251	2	A55035	cysteine-rich prot	
20	350.5	23.6	152	2	I47109	high-sulfur wool m	
21	350.5	23.6	152	2	I47112	high-sulfur wool m	
22	348	23.5	156	1	KRSHHB	keratin high-sulfu	
23	344	23.2	191	2	I46412	keratin KAP5.4 - s	
24	325	21.9	197	2	I46413	keratin KAP5.5 - s	
25	305.5	20.6	182	2	A36686	ultra-high-sulfur	
26	298.5	20.1	572	2	T29880	hypothetical prote	
27	296.5	20.0	131	1	KRGF3M	keratin high-sulfu	
28	294	19.8	728	2	T20561	hypothetical prote	
29	293	19.7	425	2	T18592	hypothetical prote	

A:Accession: A38660
A:Molecule type: DNA
A:Residues: 1-223 <WO2>
A:Cross-references: GB:M37760; NID:g200963; PIDN:AAA40107.1; PID:g200964
A:Note: this is a correction
R:Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.
J. Biol. Chem. 265, 21375-21380, 1990
A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin
A:Reference number: A38346; MUID:91065960; PMID:2250030
A:Accession: B38346
A:Molecule type: DNA
A:Residues: 1-21, 'GCGSGCGCGCGCGCGGCCCKPVCC', 22-40, 'GSS', 44-45, 'G', 47-48, 'S', 50, 'GSS', <WO2>
A:Cross-references: GB:M37759; NID:g200961; PIDN:AAA40106.1; PID:g200962
A:Note: the sequence reported in this paper has been corrected. See A38660
C:Superfamily: ultra-high-sulfur keratin

Query Match 46.7%; Score 693.5; DB 2; Length 223;
Best Local Similarity 52.7%; Pred. No. 4.4e-42;
Matches 119; Conservative 23; Mismatches 47; Indels 37; Gaps 13;

QY	4	SCGGSVCS---	DSCSGGGLQESCCRPSCC-----QTTTC--CRTTC-----CRPSC---	44
Db	2	TCCG--CSGCGGSSCG--GGC--SSCCKPVCCVPCVSCSGCGCKGCGGCGGCGGSC	57	
QY	45	--GTCSSC-----CRPSCCISSCCKPCCCLTTCCTTTC---CRPSCCISSCCRPSCCISSC	94	
Db	58	GGCKGGCGSGGCKGGCCGSSCCCKPCCCGSSCCCKPCCSSCGGSSCCGSCCKPCCCGSSC	117	
QY	95	CKPSCRTTCRPSCCISSCCRPSCCISSCCKPSCCCTTCCTTCCTTCRPSCCISSCCRPSCCISSC	154	
Db	118	CKP--CC--SSGCGSSCCGCKGCGCCGSSCCCKPCCCGSSCCCKPCCCGSSCCCKP--CCSSCGGSSCCGSSC	173	
QY	155	CKPSCCQTTCRPSCCISSCYRPPCCGCPSCCRPAACISCCCHPSCC	200	
Db	174	CKPCCGSSCCCKPCCCGSSCCCKPCCCGSSCCCKPCCCGSSCCCAPVCC	219	

RESULT 3
A38346
ultra-high-sulfur keratin 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #Text_change 24-Sep-1999
C:Accession: A38346
R:Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.
J. Biol. Chem. 265, 21375-21380, 1990
A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin
A:Reference number: A38346; MUID:91065960; PMID:2250030
A:Accession: A38346
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-230 <WO2>
A:Cross-references: GB:M37759; NID:g200961; PIDN:AAA40106.1; PID:g200962
C:Superfamily: ultra-high-sulfur keratin

```

Query Match          46.1%; Score 684; DB 2; Length 230;
Best Local Similarity 51.9%; Pred. No. 2.le-41;
Matches 124; Conservative 18; Mismatches 55; Indels 42; Gaps 14;

Qy 3 SSC--CGSVGSDSGCGGLG-----OESCCRPSCCQTTCCRTTCRPSCCISSC-- 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 SSCGGGCGSGC--CGSGGCGGCGNGCGGSGSCCKPVCC---CKPVCCCPVVCSSCGG 66

Qy 50 CRPSC-----CISSC--CKPSCGLTTCRTTCRPSCCISSC--CRPSCCISCCRPSCC 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 CGSSGCGGCGSGSGCGGSGSCCKPVC-----CCVPVCSGCGGCKPCCQSSCCRP--CC 121

Qy 101 RTTTCRPSCCIISCCRPSCCISCCRPSCCCTTCRPSCCIISCCRPSCCISCCRPSCC 160
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 122 -SSGCGSGCCQSSCCRPCCCGQSSCKPCCCGQSSCCRP--CCSSGCGSSCCQSSCCRPCC 178

Qy 161 QTTTCRPSCCIISCCRPQCCQCCRPACCISCCCHPSCCVSCRCGPFSCPTTTCRTTC 219
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

```

Db      179_QS5CCKPCCQSSCKKPVCCQSSCKPCCQSSCKP-CCSS-----GCGSSSTCCQNSC 230

RESULT 4
S60314
hair keratin cysteine rich protein - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 12-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 24-Sep-1999
C:Accession: S60314
R:Frattini, A.; Powell, B.C.; Hynd, P.I.; Keough, R.A.; Rogers, G.F.
J. Invest. Dermatol. 102, 178-185, 1994
A:Title: Dietary cysteine regulates the levels of mRNAs encoding a family of cysteine
A:Reference number: S60314; MUID:94149288; PMID:7508963
A:Accession: S60314
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-151 <FRA>
A:Cross-references: EMBL:X73462; NID:q512030; PIDN:CAA51841.1; PID:q512031
C:Superfamily: ultra-high-sulfur keratin

```

	Query Match	41.7%	Score 619;	DB 2:	Length 151:
	Best Local Similarity	61.6%	Pred. No. 5,7e-37;		
	Matches	106;	Conservative	13;	Mismatches 31; Indels 22; Gaps 3;
Qy	1	MVSSCGSVCSWQSGUJLGOESCCRPSCCQTCCRTTCCRPSCCISSCCKRPSCCISSCC	60		
Db	1	MVSSCGSVCSAOSGSLQVETCCRPSCCQTCCRTTCTYRPGSVSSCCRP	52		
Qy	61	KPSCCLTTCRTTCCRPSCCISSCCPRSCCISSCCKPSCRTTCCRPSCCISSCCKRPSCC	120		
Db	53	-----LCQPTCPPTCCISSCYRPSCCRSS-----CGSSCYRPTCSISSCKRPQCC	99		
Qy	121	ISSCKPSCCR-TTCCRPSCCISSCRPSCCISSCCKPSCQTTCRPSCCIT	171		
Db	100	OPVCCOPSGARISSCCRPSCGSGSYRPSCLRPVCNRVSHCHTTCYRPTCVI	151		

RESULT 5

146489

cysteine-rich hair keratin associated protein - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 24-Sep-1999

C:Accession: 146489; S49201

R:Powell, B.C.; Arthur, J.; Nesci, A.

Differentiation 58, 227-232, 1995

A:Title: Characterization of a gene encoding a cysteine-rich keratin associated prot.

A:Reference number: 146489; MUID:95228955; PMID:7536172

A:Accession: 146489

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-126 <POW>

A:Cross-references: EMBL:X80035; NID:9510540; PID:CAA56339.1; PID:4510541

C:Genetics:

A:Gene: KAP4L

C:Superfamily: ultra-high-sulfur keratin

	Query Match	38.4%	Score 569.5	DB 2	Length 126
	Best Local Similarity	57.7%	pred. No. 1.5e-33		
	Matches	90	Conservative 15	Mismatches 20	Indels 31
				Gaps 4	
Qy	1	MVNSCCGSGVSDQSGUGLGG-ESCGRPSCCQTTCRTTCRPPSCCIISSCCRPSCCIISSC	59		
Db		: : :			
Db	1	MVNSCCGSGVSGCGVDFQGEESCRRPS-----CGRPQCCQPSCCRPTCCIISSC	50		
Qy	60	CKPSCGLTCTGCTTCCKPSCCISSCKRPSCCIISSCKPSCCCTTCRPPSCCIISSCKPSC	119		
Db		: : :			
Db	51	CRPQCCGVCCTCCRPSCYISSCRP-----TCRPTCCRPTCCRP-----	93		
Qy	120	CISSCKRPSCGCTTCCKPSCCIISSCKRPSCCIISSC	155		
Db		: : :			
Db	94	---TSCCTCTCTCTCTCKPSCGVCCTCCRPCCSGSGCC	126		

RESULT 6

JC6547
high sulfur protein B2E - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: JC6547
R:Mitsui, S.; Ohuchi, A.: Adachi-Yamada, T.; Hotta, M.; Tsuboi, R.; Ogawa, H.
Gene 208, 123-129, 1998
A>Title: Structure and hair follicle-specific expression of genes encoding the rat high sulfur protein B2E
A:Reference number: JC6547; MJID:98201605; PMID:9524245
A:Accession: JC6547
A:Molecule type: DNA
A:Residues: 1-188 <MIT>
A:Cross-references: DDBJ:AB003753; NID:g3046870; PIDN:BAA25573.1; PID:g3046871
C:Comment: This protein is a cysteine-rich, keratin associated protein.
C:Genetics:
A:Gene: b2E
C:Superfamily: keratin high-sulfur matrix protein IIIIA
C:Keywords: hair

Query Match 37.7% Score 560; DB 2; Length 188;
Best Local Similarity 47.2%; Pred. No. 8.le-33;
Matches 94; Conservative 27; Mismatches 50; Indels 28; Gaps 7;

Qy 28 SCQTTCRC-TTCCRPSCGCISSCCRPSCCIISCKPSCCLTTCCRTCCRPSCCISCRR 86
Db ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2 ACCATSFPGFPTCTGTGGTGSGNNCCQPSCSQSSCCQPS-----CSQSACCOPSCQTQSSCCQ 56

Qy 87 PSCCIISSCCRPSCRTTCCRPSCCIISSCRPSCCIISCKKPSCC-----RTTC- 134
Db ||| ||| ||| :|||:||| :|||:||| :|||:||| :|||:||| :|||:|||
57 PTCSQSSCCOPTCSQSSCCQPSGCCOTPCCOPTCCONNSCOTSCCGTGSGQGEGGATSCR 116

Qy 135 ---CRPPSCCIISSCRPSCCIISCKPSCCOTTCCRPSCCIISSCYRPCCOCPSCCRPACCI 191
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 VRWCRPDRCVETGLPLPCCVVSCTPTCCQLLHAQAASCC-----RPSVCGSGCCRPACC- 170

Qy 192 SSCCHPSCCVSSCRCPFSC 210
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
171 CHCEPSCSKPSCSEP-SC 188

RESULT 7

S37649
high-sulfur keratin - human
C:Species: Homo sapiens (man)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
C:Accession: S37649
R:Zhumbabava, B.D.; Gening, L.V.; Gazaryan, K.G.
Mol. Biol. 26, 550-555, 1992
A>Title: Cloning and structural characterization of human hair sulfur-rich keratin gene
A:Reference number: S37649
A:Accession: S37649
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-175 <ZHU>
A:Cross-references: EMBL:X63338; NID:g311881; PIDN:CAA44938.1; PID:g311882
C:Superfamily: keratin high-sulfur matrix protein IIIA

Query Match 31.2% Score 463.5; DB 2; Length 175;
Best Local Similarity 44.3%; Pred. No. 4.le-26;
Matches 82; Conservative 19; Mismatches 51; Indels 33; Gaps 6;

Qy 33 TCCTRITCC-RPSC-----CISCCRPSCCIISSCKPSCCLTTCCTTCCRPSCC----- 80
Db ||||| | ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2 TCCQTFCGYPCSSTGTGGSQQSPSCCETSCCQPSCCETSCCQPSCCQTSFOFLASQ 61

Qy 81 ----ISSCCRPSCCIISSCKPSCCRTTCCRPSCCIGTSSCC-----RPSCCIISSCKKPSCCR 131
Db ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62 LVLDLQSCQPSCCETSQPSQSCQPSCTTSCGTCGGIGGGIYGVEGSSCAVST-----R 114

Qy 132 TTCCRPSCCIISSCRPSCCIISCKPSCCQTTCCRPSCCIISSCYRPQCQPSCCRPACCI 191
Db ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
115 IRWCPRDRCVBGTCLLPCCVVVSCHTPTCCQLLHAEASGC-----RPSYCGSGCCRPVCDC 169

Qy 192 SSCCH 196
II I
Db 170 YSCSH 174

RESULT 8
S37650
high-sulfur keratin - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
C:Accession: S37650
R:Zhumbabaeva, B.D.; Gening, L.V.; Gazaryan, K.G.
Mol. Biol. 26, 550-555, 1992
A:Title: Cloning and structural characterization of human hair sulfur-rich keratin gene
A:Reference number: S37649
A:Accession: S37650
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <ZHU>
A:Cross-references: EMBL:X63337; NID:g311879; PIDN:CAA44937.1; PID:g311880
C:Superfamily: keratin high-sulfur matrix protein IIIA

Query Match 30.4%; Score 451; DB 2; Length 177;
Best Local Similarity 44.0%; Pred. No. 3e-25;
Matches 80; Conservative 16; Mismatches 70; Indels 16; Gaps 5;

Qy 28 SCQQTTCRTTCCRPSS-CGISSCCRPSGCCISSCKPSCCLTTCCRTTCCCKPSCCG-ISSCC 85
||||| : | ||||| ||||| : ||||| ||||| : ||||| ||||| : |||||
Db 2 ACCOTSPGGPPSGSTSGTGSSCCQPSCCETSSCCPRCCTSCCQPSCTSPFGFSES 61

Qy 86 RPSCGISSSCKPSCCLRTTCCRPSCCISSSCCRPSCCI-----SSCCRPSCCLTTCCRP 137
| ||||| ||||| ||||| ||||| : ||||| ||||| : ||||| ||||| : |||||
Db 62 TGGTCUSSCCCPSCCETSCCQPSCYQTSSCGTCGGIGGYGOGSSCAVSTRIRMCRLP 121

Qy 138 SCCTSSCCRPSCCTSCCKPSCCQTTCCRPSCCTSSCYRPCCQPCPCRPAACIISSCGHP 197
: | ||||| ||||| ||||| ||||| : ||||| ||||| : ||||| ||||| : |||||
Db 122 DCRVEGTPLPCCVSVCTPSCCOLHIAASCC-----RPSYCGSCCRPVCC-CYCSEP 175

Qy 198 SC 199
: |
Db 176 TC 177

RESULT 9
KRSHHD
keratin high-sulfur matrix protein B2d - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 22-Jun-1999
C:Accession: S07911
R:Powell, B.C.; Sleight, M.J.; Ward, K.A.; Rogers, G.E.
Nucleic Acids Res. 11, 5327-5346, 1983
A:Title: Mammalian keratin gene families: organisation of genes coding for the B2 high-sulfur matrix protein IIIA
A:Reference number: S07349; MUID:83299218; PMID:6193483
A:Accession: S07911
A:Molecule type: DNA
A:Residues: 1-182 <POW>
A:Cross-references: FMBI:x01610; NID:g1205; PIDN:CAA25759.1; PID:g1298
C:Superfamily: keratin high-sulfur matrix protein IIIA
C:Keywords: duplication; hair
F:2-182/Product: Keratin high-sulfur matrix protein B2d #status predicted <MAT>
F:27-36,37-46,47-56,57-66,67-76,77-86/Region: duplication

Query Match 29.2%; Score 434; DB 1; Length 182;
Best Local Similarity 37.1%; Pred. No. 4.7e-24;
Matches 76; Conservative 27; Mismatches 54; Indels 48; Gaps 7;

Qy 4 SCCG-SVCSDSGGLGQESCCRPSCCQTTCCRTTCCRPSCCISSSCCRPSCCG-ISSCCPK 62
||||| : | ||||| ||||| : ||||| ||||| : ||||| ||||| : |||||
Db 2 ACCSTSPGPHPTSTVG----GTCGSNFCQPCQVCQTSCTSCPSTOTSCQPSITGTSQCP 57

Qy 63 SCCLTTCRTTCCRPSCCISSSCCRPSCCISSCCKPSCC-----CRTTC 104
| : ||||| : ||||| ||||| : ||||| ||||| : |||||

Db 58 -----TSIQTSCCQPISIQTSCCQPTCLQTSGETGCGIGSGYGVGGSSGAVSSRTKW 112
Qy 105 CRPSCCISCCRPSCCISCCCKPSCC-----RTTCRPSCCISCCRPSCCISCCCKPSC 159
Db 113 CRPDRVETSLPCCVCCVSPSCCQLYAAQASCCRPSCYCGQCCCRPAC-----C 163
Qy 160 QCTTCRPSCCISCCRPSCCQPCSC 184
Db 164 CQPTCIEPVC-----EPTCCEPTC 182

RESULT 10
147105
high-sulfur wool matrix protein B2A - sheep
A:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 15-Sep-2000
C:Accession: I47105
R:Rogers, G.R.; Hickford, J.G.; Bickerstaffe, R.
Anim. Genet. 25, 407-415, 1994
A:Title: Polymorphism in two genes for B2 high sulfur proteins of wool.
A:Reference number: I47105; MUID:95209146; PMID:7695121
A:Accession: I47105
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-182 <ROG>
A:Cross-references: GB:L33885; NID:g499871; PIDN:AA01444.1; PID:g940355
C:Superfamily: keratin high-sulfur matrix protein IIIA

Query Match 29.1%; Score 432; DB 2; Length 182;
Best Local Similarity 40.7%; Pred. No. 6.5e-24;
Matches 74; Conservative 25; Mismatches 57; Indels 26; Gaps 5;

Qy 43 SCQISCCRPSCCISCCCKPSCC-----RTTCRPSCCISCCRPSCCISCCCKPSCCR 101
Db 2 ACCSTFCGFPICSTGGTGGSSPCQTCCTCCQTSQTSQTSQTSQTSQTSQTSQTSQTSQ 61

Qy 102 TTCCRPSCCISCCRPSCCISCCCKPSCC-----RTTCRPSCCISCCCKPSCCISCCCKPSCC 143
Db 62 TSCCQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQTSQ 121

Qy 144 CRPSCCISCCCKPSCC-----RTTCRPSCCISCCRPSCCISCCCKPSCCISCCCKPSCC 197
Db 122 TSLPCCVCCVTPPSCCQLYAAQASCCRPSCYCGQCCCRPACCCQPTCTEPV-CEPSCCEP 180

Qy 198 SC 199
Db 181 TC 182

RESULT 11
S18946
ultra high-sulfur keratin 1 - human
A:Alternate names: UHS keratin; ultra high-sulfur matrix protein
C:Species: Homo sapiens (man)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: S18946; B36686
R:Drabant, B.; Donecke, D.
submitted to the EMBL Data Library, December 1991
A:Description: Nucleotide sequence of a human high-sulfur keratin cDNA.
A:Reference number: S18946
A:Accession: S18946
A:Molecule type: mRNA
A:Residues: 1-169 <DRA>
A:Cross-references: EMBL:X63755; NID:g32471; PIDN:CAA5283.1; PID:g32472
R:MacKinnon, P.J.; Powell, B.C.; Rogers, G.E.
J. Cell Biol. 111, 2587-2600, 1990
A:Title: Structure and expression of genes for a class of cysteine-rich proteins of the
A:Reference number: A36686; MUID:91115951; PMID:1703541
A:Accession: B36686
A:Molecule type: DNA
A:Residues: 1-39, 'Y', 41-169 <MAC>
A:Cross-references: GB:X55293; NID:g34078; PIDN:CAA39005.1; PID:g34079
C:Genetics:

A:Gene: GDB:KRN1
A:Cross-references: GDB:I25257; OMIM:148021
A:Map position: 11q13-11q13
C:Superfamily: ultra-high-sulfur keratin
C:Keywords: hair; tandem repeat
F:7-15/Region: Ser-rich nonapeptide repeat
F:59-68/Region: Gly-rich decapeptide repeat
F:69-78/Region: Gly-rich decapeptide repeat
F:79-88/Region: Cys-rich decapeptide repeat
F:89-97/Region: Ser-rich nonapeptide repeat
F:98-107/Region: Cys-rich decapeptide repeat
F:108-117/Region: Cys-rich decapeptide repeat
F:118-126/Region: Ser-rich nonapeptide repeat
F:127-136/Region: Cys-rich decapeptide repeat
F:137-145/Region: Ser-rich nonapeptide repeat
F:146-155/Region: Cys-rich decapeptide repeat
F:156-165/Region: Cys-rich decapeptide repeat

Query Match 29.1%; Score 431.5; DB 1; Length 169;
Best Local Similarity 46.9%; Pred. No. 6.8e-24;
Matches 82; Conservative 13; Mismatches 55; Indels 25; Gaps 8;

Qy 3 SSCGSGVCS--DQSCGGGIGQESG-----CRPSCCQTTCCRTTCRPSCCISCCCKPSC 54
Db 7 SGGCGSGCGGCDSSCG-----SCGSGCGCGCPSCCAPVC---CCKPVC---CCVPAC 52

Qy 55 C1SSCCRPSCCCTTCCTTCRPSCCISCCCKPSCCISCCCKPSCCCTTCRPSCCISCCCKPSC 114
Db 53 SCSSGGRGCGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 111

Qy 115 CRPSCCISCCCKPSCCCTTCRPSCCISCCCKPSCCISCCCKPSCCCTTCRPSCCISCCCKPSC 169
Db 112 CKP-CCSSSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSG 164

RESULT 12
KRSHA
keratin high-sulfur matrix protein B2A - sheep
A:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 24-Apr-1984 #sequence_revision 12-Apr-1996 #text_change 22-Jun-1999
C:Accession: S07910; A02837
R:Powell, B.C.; Steigh, M.J.; Ward, K.A.; Rogers, G.E.
Nucleic Acids Res. 11, 5327-5346, 1983
A:Title: Mammalian keratin gene families: organisation of genes coding for the B2 high
A:Reference number: S07349; MUID:83299218; PMID:6193483
A:Accession: S07910
A:Molecule type: DNA
A:Residues: 1-172 <POW>
A:Cross-references: EMBL:X01610; NID:g1295; PIDN:CAA25757.1; PID:g1296
R:Elleman, T.C.
Biochem. J. 130, 833-845, 1972
A:Title: The amino acid sequence of protein SCMK-B2A from the high-sulphur fraction of
A:Reference number: A02837; MUID:73224964; PMID:4679226
A:Accession: A02837
A:Molecule type: protein
A:Residues: 2-172 <ELL>
A:Experimental source: Lincoln wool
A:Note: a minor component has 13-Thr, 23-Asn, and 24-Phe
C:Comment: The keratin products of mammalian epidermal derivatives such as wool and hair
fur and high-tyrosine keratins, having molecular weights of 6-20 kilodaltons, whereas
C:Superfamily: keratin high-sulfur matrix protein IIIA
C:Keywords: acetylated amino end; blocked amino end; duplication; hair
F:2-172/Product: keratin high-sulfur matrix protein B2A #status experimental <MAT>
F:2-76/Region: 10-residue repeats
F:2-7/Modified site: acetylated amino end (Ala) (in mature form) #status experimental

Query Match 27.5%; Score 408.5; DB 1; Length 172;
Best Local Similarity 37.6%; Pred. No. 2.7e-22;
Matches 76; Conservative 21; Mismatches 58; Indels 47; Gaps 7;

Qy 4 SGGCGSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62
Db 2 ACCSTFCGFPICSTGGTGGSSPCQTCCTCCQTSQTSQTSQTSQTSQTSQTSQTSQTSQ 57

19

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 15:10:00 ; Search time 224 Seconds
(without alignments)
7212.667 Million cell updates/sec

Title: US-09-874-062-2
Perfect score: 1101
Sequence: 1 ctgggaaccaccagagacc.....caaaaaaaaaaaaaaaaaa 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1101	100.0	1101	10	US-09-874-062-2
2	198	18.0	461	10	US-09-864-761-10890
3	187	17.0	2108	10	US-09-962-832-225
4	167.4	15.2	8705	9	US-10-291-230-14
5	165.4	15.0	489	10	US-09-864-761-27528
6	160.2	14.6	478	10	US-09-864-761-6314
7	146.6	13.3	403	10	US-09-864-761-22764
8	142	12.9	11872	9	US-10-092-154-1390
9	142	12.9	11872	10	US-09-764-847-1390
10	137.4	12.5	463	10	US-09-864-761-5295
11	136.6	12.4	1974	10	US-09-864-761-4747
12	132.2	12.0	464	10	US-09-864-761-6005
13	129	11.7	1101	10	US-09-874-062-2
14	126.2	11.5	696	10	US-09-864-761-21486
15	122.6	11.1	364	10	US-09-864-761-22070
16	119.2	10.8	256	10	US-09-783-590-3765
17	118.2	10.7	422	9	US-09-854-133-337
18	118.2	10.7	422	10	US-09-738-973-337
19	116.8	10.6	442	10	US-09-864-761-11569

ALIGNMENTS

RESULT 1
US-09-874-062-2
; Sequence 2, Application US/09874062
; Patent No. US20020081607A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Four Disulfide Core Domain-Containing (FDCD) Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT033P1
; CURRENT APPLICATION NUMBER: US/09/874,062
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US00/32462
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/168,229
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-874-062-2

Query Match	100.0%	Score 1101	DB 10	Length 1101
Best Local Similarity	100.0%	Pred. No. 2.8e-260	Mismatches 0	Indels 0
Matches 1101	Conservative	0	0	Gaps 0
Qy	1	CTGGGAACCCACCCAGAACCTCCACCCCTCTCAGCGCATGGTCAGCTCCTGCTGTGGCTCT 60		
Db	1	CTGGGAACCCACCCAGAACCTCCACCCCTCTCAGCGCATGGTCAGCTCCTGCTGTGGCTCT 60		
Qy	61	GTCTGCTCTGACGAGAGTGTTGGTCAAGGTCGCGCCAGGAGAGTGCTGCGGCGGAGC 120		
Db	61	GTCTGCTCTGACGAGAGTGTTGGTCAAGGTCGCGCCAGGAGAGTGCTGCGGCGGAGC 120		
Qy	121	TGCTGCCAGACCCACCTCTCTGAGGACACCTGCTGCGCCGCCAGCTGCTGCAATTTCCAGT 180		
Db	121	TGCTGCCAGACCCACCTCTCTGAGGACACCTGCTGCGCCGCCAGCTGCTGCAATTTCCAGT 180		
Qy	181	TGCTGCAAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGCAACACC 240		
Db	181	TGCTGCAAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGCAACACC 240		

Sequence 5592, Ap
Sequence 6987, Ap
Sequence 4, Appli
Sequence 510, App
Sequence 157, App
Sequence 12, Appli
Sequence 5097, Ap
Sequence 21820, A
Sequence 37, Appli
Sequence 31, Appli
Sequence 225, App
Sequence 1601, Ap
Sequence 62, Appli
Sequence 66, Appli
Sequence 10, Appli
Sequence 11, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 23031, A
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 105, App

20 114.4 10.4 573 10 US-09-864-761-5592
21 112.6 10.2 22585 9 US-09-764-891-6987
22 111 10.1 1852 10 US-09-969-852-4
23 110.2 10.0 1562 9 US-10-037-270-510
24 105.6 9.6 2313 9 US-09-854-133-157
25 105.6 9.6 2313 10 US-09-738-973-157
26 105.6 9.6 2314 9 US-09-764-868-12
27 103 9.4 1967 10 US-09-864-761-5097
28 102.8 9.3 648 10 US-09-864-761-21820
29 102.2 9.3 3015 9 US-09-373-658-37
30 100.2 9.1 3331 9 US-09-373-658-31
31 99.8 9.1 2108 10 US-09-962-832-225
32 99.2 9.0 14800 10 US-09-954-456-1601
33 98.6 9.0 1649 9 US-09-789-390-66
34 98.6 9.0 1787 9 US-09-789-390-66
35 98.6 9.0 3822 9 US-09-789-390-10
36 98.6 9.0 3999 9 US-09-789-390-6
37 97 8.8 4133 9 US-09-922-469-3
38 95.6 8.7 657 10 US-09-922-469-3
39 95.6 8.7 657 10 US-09-922-480-3
40 95.6 8.7 657 10 US-09-923-236-3
41 93 8.4 256 10 US-09-864-761-23031
42 92.4 8.4 1008 9 US-09-922-469-1
43 92.4 8.4 1008 10 US-09-922-480-1
44 92.4 8.4 1008 10 US-09-923-236-1
45 92.4 8.4 1325 9 US-09-746-783-105

;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonam Sequence Listing Engine vers. 1.1
;; SEQ ID NO 22764
;; LENGTH: 403
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC006070.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
;; OTHER INFORMATION: EST_HUMAN HIT: BE250684.1, EVALU 4.00e-07
;; OTHER INFORMATION: NT HIT: q17657286, EVALU 8.00e-20
;; OTHER INFORMATION: SWISSPROT HIT: P02438, EVALU 7.50e-02
US-09-864-761-22764

Query Match 13.3%; Score 146.6; DB 10; Length 403;
Best Local Similarity 65.0%; Pred. No. 3.2e-26;
Matches 266; Conservative 0; Mismatches 134; Indels 9; Gaps 3;
QY 175 TCCAGTCTGCGAGCCCTTCTGCTATCTCCAGCTGCTGCAACCCAGCTGCTGCGCTG 234
DB 400 TCCCTGCTGCTGAGCCCTTACCTGCTGCGAGGACCACTGCTGGAAGCCCA---CCACTGTG 344
QY 235 ACCAGCTGCTGAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTCCAGTTGCTGCGAG 294
DB 343 ACCAGCTGCTGAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTCCAGTTGCTGCGAG 284
QY 295 CTTCTGCTGCTATCTCCAGCTGCTGCAAAACCCAGCTGCTGAGGACCACTGCTGCGCG 354
DB 283 CCT---TGCTGCCACCACTTGTGCTGCAAAACACCTGCTGAGGACCACTGCTGCGCG 227
QY 355 CCCAGCTGCTGATTTCCAGTTGCTGCGAGGCTTCTGCTATCTCCAGTTGCTGCAAA 414
DB 226 CCCA---CTGTGTGACAGCTGCTGCGAGGCTTCTGCTGCGAGGACCACTGCTGCGCG 170
QY 415 CCCAGCTGCTGAGGACCACTGCTGCGCGCCAGCTGCTGCAATTTCTAGTTGCTGCGAG 474
DB 169 CCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 110
QY 475 CTTCTGCTGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 534
DB 109 CCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 50
QY 535 CCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 583
DB 49 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1

RESULT 8
US-10-092-154-1390
;; Sequence 1390, Application US/10092154
;; Publication No. US20030054375A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC009C1
;; CURRENT APPLICATION NUMBER: US/10/092,154
;; CURRENT FILING DATE: 2002-03-07
;; NUMBER OF SEQ ID NOS: 2003
;; Prior application removed - See File Wrapper or Palm
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1390
;; LENGTH: 11872
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-092-154-1390

Query Match 12.9%; Score 142; DB 9; Length 11872;
Best Local Similarity 60.8%; Pred. No. 4.1e-24;
Matches 305; Conservative 0; Mismatches 185; Indels 12; Gaps 4;
QY 222 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 281
DB 5348 CAGAAGCCCACTTCCATCCCTGACACCATGACCCCACTGTTGCTCCCTTCTGCTGCTG 5407
QY 282 CAGTTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 341
DB 5408 TACCTGCTGAGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5464
QY 342 CACCTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 401
DB 5465 ACCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5521
QY 402 CAGCTGCTGCAAAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 461
DB 5522 AACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5578
QY 462 TAGTTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 521
DB 5579 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5638
QY 522 CACCTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
DB 5639 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5698
QY 582 CTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 641
DB 5699 CACAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5758
QY 642 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701
DB 5759 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5815
QY 702 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 723
DB 5816 CACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5837

RESULT 9

US-09-764-847-1390
;; Sequence 1390, Application US/09764847
;; Patent No. US20020132767A1
;; GENERAL INFORMATION:
;; APPLICANT: Rosen et al.
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: PC009
;; CURRENT APPLICATION NUMBER: US/09/764,847
;; CURRENT FILING DATE: 2001-01-17
;; Prior application data removed - consult PALM or file wrapper
;; NUMBER OF SEQ ID NOS: 2003
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 1390
;; LENGTH: 11872
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-764-847-1390

Query Match 12.9%; Score 142; DB 10; Length 11872;
Best Local Similarity 60.8%; Pred. No. 4.1e-24;
Matches 305; Conservative 0; Mismatches 185; Indels 12; Gaps 4;
QY 222 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 281
DB 5348 CAGAAGCCCACTTCCATCCCTGACACCATGACCCCACTGTTGCTCCCTTCTGCTGCTG 5407
QY 282 CAGTTGCTGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 341
DB 5408 TACCTGCTGAGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5464


```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21486
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003958.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: NT HIT: L33886.1, EVALUATION = 9.60e-02
; OTHER INFORMATION: EST_HUMAN HIT: A1907096.1, EVALUATION = 1.00e-116
; OTHER INFORMATION: SWISSPROT HIT: P08131, EVALUATION = 9.00e-16
US-09-864-761-21486

Query Match 11.5%; Score 126.2; DB 10; Length 696;
Best Local Similarity 52.4%; Pred. No. 4.8e-21;
Matches 312; Conservative 0; Mismatches 268; Indels 15; Gaps 1;

QY 119 GCTGCTGCCAGACACCTGCTGCAGGACCACCTGCTGCCGCCGCCAGCTGCTGCATTTCCTCA 178
Db 1 GCTGCTGCCAGCCTGTAGGCTCTGAAGCCACTTCTGCGCAACACAGCTCTGCTGTGCGCCA 60
QY 179 GTTCTGTCGAGCCCTTCCTGCTGTATCTCCACCTGCTGCAACCCAGCTGCTGCCTGACCA 238
Db 61 CTTCCTGCCAGCCTGTCTCTGCAAAATCCACCTGCTGCCAGCCAGTTCCTGTGTAGGCCCA 120
QY 239 CTTGCTGAGGACACCTGCTGCCGCCGCCAGCTGCTGCATTTCCTGCTGCAGGCGCTT 298
Db 121 GCTGCTGTTGACCTGCTGTCGACCCCTGCTAGTCTGCGCAACCTGCTGCTGTGAGCCTT 180
QY 299 CTTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAGGACACCTGCTGCGCGGCCCA 358
Db 181 CTTGCTGTGAGCGGCTGTGCGCGGACACCTGCTGCTGTGACCAAGCTGCTGCGCGGCTG 240

359 GCTGCTGCATTTCAGCTTGTGTCAGGCGCTTCCTGCTGTATCTCCAGCTGCTGCAAAACCA 418
Db 241 TCTGCTGTGAGCCAG-----CCCTTGTGAGCAAGTTGCTCAGAGTCTA 285
QY 419 GCTGCTGCAGGACACCTGCTGCCGCCGCCAGCTGCTGCATTTCCTGCTGCAAGGCTTT 478
Db 286 GCATCTGCCAGCCAGCTTACCTGTGTGCTGCTGCTGAGCCAGTTTGTGCTGCTGCTG 345
QY 479 CTTGCTGTATCTTAGCTTGTGCAAAACCCAGCTGCTGCCAGACACCTGCTGCTGCTGCA 538
Db 346 TCTGCTGTGTTTCAAGCTTGTGCGAGCCAGCTTCTGTCCCGAGCATTGTCGCAAGGCTTT 405
QY 539 GCTGCTGTATCTCAGCTGTACAGGCCGCCAGCTGCTGCCAGCCCTCTGCTGCTGCTGCTG 598
Db 406 CTTGCTGTGCTCAGCTGTGCAACCCAGCTGCTGCTGAGCCAGCCCTGCTGCTGCTGCTG 465
QY 599 CTTGCTGCATTTCAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 658
Db 466 CTGCTGCTGTGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525
QY 659 TCAGCTGCCGCCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 713
Db 526 TCTGCCCTGAGCCAGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 580

RESULT 15
US-09-864-761-22070/C
; Sequence 22070, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
```


: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
: SEQ ID NO 22070
: LENGTH: 364
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC006070.1
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
: OTHER INFORMATION: SWISSPROT HIT: P45931, EVALUE 1.50e+00
: OTHER INFORMATION: EST_HUMAN HIT: BE250684.1, EVALUE 2.00e-05
: OTHER INFORMATION: NT HIT: M27685.1, EVALUE 2.00e-17
US-09-864-761-22070

Query Match		11.1%	Score 122.6;	DB 10;	Length 364;	
Best Local Similarity		64.0%	Pred. No. 2.4e-20;			
Matches 236;		Conservative	0;	Mismatches 124;	Indels 9;	Gaps 3;
QY	118	AGCTGCTGCCAGACCACTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCATTTC	177			
Db	364	AGAAGCCACCCCTCCACCCCTGACACATGACCCACTGTTGCCCTGCTGCAGCCT	305			
QY	178	AGTTGCTGCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAGCCAGCTGCTGCCTGACC	237			
Db	304	ACCTGCTGCAGGACCACTTCTGCAGGACTACCTGTGGCAGCCCA--CCATTGTGACC	248			
QY	238	ACCTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCATTTCCAGTTGCTGCAGGCCT	297			
Db	247	ACCTGCAGCAGCACACCCTGCTGCCAGCCCTCCTGCTGTGTGTCAGCTGCTGCCAGCCT	188			
QY	298	TCCTGCTGTATCTCCAGCTGCTGCAACCCAGCTGCTGCAGGACCACTGCTGCCGCC	357			
Db	187	TACTGC---CACCCACTTCTGTCAAAACACCTGCTGCAGGACCACTGCTGCCAGCCC	131			
QY	358	AGCTGCTGCATTTCACATTTGCTGCAGGCTTTCCTGCTGTATCTCCAGCTGCTGCAACCC	417			
Db	130	A---CCTGTGTGACCACTGCTGCCAGCCTTCTGCTGCAGCAGCACACCCTGCTACCAGCCC	74			
QY	418	AGCTGCTGCAGGACCACTGCTGCCGCCAGCTGCTGCATTTCTAGTTGCTGCAGGCCT	477			
Db	73	ATCTGCTGTGGTCCAGCTGCTGTGGCCAAACCAAGCTGTGGTCCAGCTGTGGCCAGAGC	14			
QY	478	TCCTGCTGT 486				
Db	13	AGCTCCTGT 5				

Search completed: June 18, 2003, 17:55:55
Job time : 231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 10:15:25 ; Search time 70 seconds
(without alignments)
4823.587 Million cell updates/sec

Title: US-09-874-062-2

Perfect score: 1101

Sequence: 1 ctgggaaccaccagacc.....caaaaaaaaaaaaaaaaaaa 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A-COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B-COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A-COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B-COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PCRUS-COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	252	22.9	3489	2	US-08-728-323A-1
C 2	252	22.9	3489	4	US-09-298-568-1
C 3	252	22.9	32207	2	US-08-770-379-20
C 4	252	22.9	32207	4	US-08-757-669A-20
C 5	252	22.9	32207	4	US-09-230-371A-20
C 6	167.4	15.2	1926	4	US-09-249-585A-2
C 7	167.4	15.2	2580	3	US-09-050-863-2
C 8	167.4	15.2	2580	4	US-09-359-081-2
C 9	167.4	15.2	5452	2	US-09-130-114-1
C 10	167.4	15.2	9600	4	US-08-910-647-1
C 11	167.4	15.2	9600	4	US-09-620-925-1
C 12	167.4	15.2	10596	1	US-07-884-811-15
C 13	167.4	15.2	10596	1	US-07-885-971-15
C 14	167.4	15.2	10596	1	US-08-087-783A-15
C 15	167.4	15.2	10596	1	US-08-194-088B-15
C 16	167.4	15.2	10596	2	US-08-194-087-15
C 17	167.4	15.2	10596	5	PCT-US93-04648-15
C 18	151	13.7	2824	2	US-09-010-928B-3
C 19	139.4	12.7	2830	2	US-09-010-928B-1
C 20	134.6	12.2	3489	2	US-08-728-323A-1
C 21	134.6	12.2	3489	4	US-09-298-568-1
C 22	134.6	12.2	32207	2	US-08-770-379-20
C 23	134.6	12.2	32207	4	US-08-757-669A-20
C 24	134.6	12.2	32207	4	US-09-230-371A-20
C 25	128.6	11.7	2793	1	US-08-209-747-1
C 26	128.6	11.7	2793	1	US-08-458-298-1
C 27	126.6	11.5	9551	1	US-08-056-200-93

C 28	126.6	11.5	9551	2	US-08-800-644-93	Sequence 93, Appli
C 29	122.4	11.1	543	6	5273901-6	Patent No. 5273901
C 30	112.8	10.2	2338	1	US-08-425-069-1	Sequence 1, Appli
C 31	112.8	10.2	2338	2	US-08-317-844B-1	Sequence 1, Appli
C 32	112.4	10.2	533	6	5482709-5	Patent No. 5482709
C 33	111.2	10.1	2830	2	US-09-010-928B-1	Sequence 1, Appli
C 34	106.2	9.6	1995	1	US-08-425-069-3	Sequence 3, Appli
C 35	106.2	9.6	1995	2	US-08-317-844B-3	Sequence 3, Appli
C 36	105.6	9.6	2313	4	US-09-370-838-157	Sequence 157, App
C 37	104.2	9.5	543	6	5273901-6	Patent No. 5273901
C 38	101.4	9.2	397	3	US-09-253-691-3	Sequence 3, Appli
C 39	100.2	9.1	2214	3	US-08-864-038A-1	Sequence 1, Appli
C 40	100.2	9.1	3331	3	US-08-864-038A-2	Sequence 2, Appli
C 41	100.2	9.1	3331	3	US-08-864-038A-4	Sequence 4, Appli
C 42	99	9.0	3095	6	5231168-1	Patent No. 5231168
C 43	95.6	8.7	657	4	US-09-527-345-3	Sequence 4, Appli
C 44	93.4	8.5	2793	1	US-08-209-747-1	Sequence 1, Appli
C 45	93.4	8.5	2793	1	US-08-458-298-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-08-728-323A-1/c

: Sequence 1, Application US/08728323A

: Patent No. 5948676

: GENERAL INFORMATION:

: APPLICANT: Chang, Yuan

: APPLICANT: Bohenzky, Roy A.

: APPLICANT: Russo, James J.

: APPLICANT: Edelman, Isidore S.

: APPLICANT: Moore, Patrick S.

: TITLE OF INVENTION: Immediate Early Protein From Kaposi's

: TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA

: TITLE OF INVENTION: Encoding Same And Uses Thereof

: NUMBER OF SEQUENCES: 21

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Cooper & Dunham LLP

: STREET: 1185 Avenue of the Americas

: CITY: New York

: STATE: New York

: COUNTRY: U.S.A.

: ZIP: 10036

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/728,323A

: FILING DATE:

: CLASSIFICATION: 435

: ATTORNEY/AGENT INFORMATION:

: NAME: White, John P.

: REGISTRATION NUMBER: 28,678

: REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 212-278-0400

: TELEFAX: 212-391-0525

: INFORMATION FOR SEQ ID NO: 1:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 3489 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: Single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

: FEATURE:

: NAME/KEY: CUS

: LOCATION: 1..3489

US-08-728-323A-1

Query Match 22.9% Score 252; DB 2; Length 3489;

```
Best Local Similarity 59.3%; Pred. No. 1.7e-42;
Matches 429; Conservative 0; Mismatches 295; Indels 0; Gaps 0;
QY 42 CAGTCTCCTGCTGTGGCTCTGCTCTGTGACAGAGCTGTGGTCAAGGTCTGCGCCAGGA 101
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2136 CTGCTCATCCTGCTGCTCTGCTCATCTGCTGCTGCTCATCTGCTGCTCATC 2077
QY 102 GAGCTGTGCGCGCCAGCTGCTGCCAGACACCTGCTGCGAGGACCACTGCTGCGCGCC 161
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2076 CTGCTGTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 2017
QY 162 CAGTGTGCTGATTCACGCTGCTGCGAGGCTTCCTGCTGTATCTCCAGCTGCTGCAACG 221
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2016 CTGCTGCTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1957
QY 222 CAGTGTGCTGCTGACACCTGCTGCCAGGACCACTGCTGCCGCGCCAGCTGCTGCAATTC 281
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1956 CTGCTGCTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1897
QY 282 CAGTGTGCTGCAACCCAGCTGCTGCCAGGACCACTGCTGCCGCGCCAGCTGCTGCAATTC 341
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1896 CTGCTGCTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1837
QY 342 CAGTGTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 401
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1836 CTGCTGCTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1777
QY 402 CAGTGTGCTGCAACCCAGCTGCTGCCAGGACCACTGCTGCCGCGCCAGCTGCTGCAATTC 461
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1776 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1717
QY 462 TAGTGTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1657
QY 522 CAGTGTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1656 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1597
QY 582 CTCTGCTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1596 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1537
QY 642 CAGTGTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 701
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1536 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477
QY 702 CATCTGCTGCGCGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1476 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1417
QY 762 CTCT 765
  |||
Db 1416 CTGT 1413
```

RESULT 2

```
US-09-298-568-1/c
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kave, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1
```

Query Match 22.9%; Score 252; DB 4; Length 3489;

Best Local Similarity 59.3%; Pred. No. 1.7e-42;

Matches 429; Conservative 0; Mismatches 295; Indels 0; Gaps 0;

QY 42 CAGTCTCCTGCTGTGGCTCTGCTCTGTGACAGAGCTGTGGTCAAGGTCTGCGCCAGGA 101

Db 2136 CTGCTCATCCTGCTGCTCTGCTCATCTGCTGCTGCTCATCTGCTGCTCATC 2077

QY 102 GAGCTGTGCGCGCCAGCTGCTGCCAGACACCTGCTGCGAGGACCACTGCTGCGCGCC 161

Db 2076 CTGCTGTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 2017

QY 162 CAGTGTGCTGATTCACGCTGCTGCGAGGCTTCCTGCTGTATCTCCAGCTGCTGCAACG 221

Db 2016 CTGCTGCTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1957

QY 222 CAGTGTGCTGCTGACACCTGCTGCCAGGACCACTGCTGCCGCGCCAGCTGCTGCAATTC 281

Db 1956 CTGCTGCTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1897

QY 282 CAGTGTGCTGCAACCCAGCTGCTGCCAGGACCACTGCTGCCGCGCCAGCTGCTGCAAGAC 341

Db 1896 CTGCTGCTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTCATC 1837

QY 342 CAGTGTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 401

Db 1836 CTGCTGCTGCTCATCCTGCTGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1777

QY 402 CAGTGTGCTGCAACCCAGCTGCTGCCAGGACCACTGCTGCCGCGCCAGCTGCTGCAATTC 461

Db 1776 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1717

QY 462 TAGTGTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 521

Db 1716 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1657

QY 522 CAGTGTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581

Db 1656 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1597

QY 582 CTCTGCTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 641

Db 1596 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1537

QY 642 CAGTGTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 701

Db 1536 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1477

QY 702 CATCTGCTGCGCGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761

Db 1476 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1417

QY 762 CTCT 765

Db 1416 CTGT 1413

RESULT 3

```
US-08-770-379-20
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohlenky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
```

```

: TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
: TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/770,379
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 52342
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 32207 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-770-379-20

```

[illegible][illegible]

	Query Match	22.9%	Score: 252;	DB 4;	Length 32207;
	Best Local Similarity	59.3%	Pred. No. 3.2e-42;		
	Matches 429;	Conservative	0;	Mismatches 295;	Indels 0; Gaps 0;
QY	42	CAGCTGCTGCTGGGCTCTGCTGCTGACCAAGACTGTGGTCAAGGTCGTGGCCAGGA	101		
DB	19861	CTGCTCATCTGCTGTCTGCTGCTCATCTGCTGCTCATCTGCTGCTGCTGCTCATC	19920		
QY	102	GAGTGTGCTGGCGGCGCCAGCTGCTGGCAGACCACTGCTGCAGACCACTGCTGCTGGCC	161		


```

: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1926)
: OTHER INFORMATION: coding strand of EBNA-1 DNA
: US-09-249-585A-2

Query Match      15.2%; Score 167.4; DB 4; Length 1926;
Best Local Similarity 53.0%; Pred. No. 2e-25;
Matches 382; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

Qy 7 ACCACCCAGAACCTCCACCTCTGACGCCATGTCAGCTCCTGCTGCTGCTGCTGCTG 66
Db 1014 ACCCGGCTCCACTGCTCTGACCGCGGCTCCACCTCCTGCTGCTGCTGCTGCTG 955
Qy 67 TCTGACACAGAGCTGTGCTCAAGCTCTCGGCAGAGAGCTGCTGCCGCCAGCTGCTG 126
Db 954 TCTGCCCCCTCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 895
Qy 127 CAGACCACTGCTGCAGGACACCTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTG 186
Db 894 TCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 835
Qy 187 AGGCTTTCTGCTGCTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGACACCTGCTG 246
Db 834 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
Qy 247 AGGACCACTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
Db 774 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 718
Qy 307 ATCTCCAGCTGCTGCAAAACCCAGCTGCTGCAAGACCACTGCTGCTGCTGCTGCTGCT 366
Db 717 CCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658
Qy 367 ATTTCCAGTTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 426
Db 657 TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
Qy 427 AGGACCACTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
Db 597 TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
Qy 487 ATCTCTAGCTGCTGCAAAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 546
Db 537 TCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 478
Qy 547 ATCTCCAGCTGCTACAGGCCCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606
Db 477 TGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
Qy 607 ATTTCTAGTTGCTGTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666
Db 417 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
Qy 667 CCGACCACTGCTGTAGAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726
Db 357 TCCTGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 298
Qy 727 T 727
Db 297 T 297

RESULT 7
US-09-050-863-2/c
: Sequence 2, Application US/09050863
: Patent No. 6114111
: GENERAL INFORMATION:
: APPLICANT: Lao, Ying
: APPLICANT: Hiang, Betty
: APPLICANT: Payan, Don
: TITLE OF INVENTION: Mammalian Protein Interaction Cloning
: TITLE OF INVENTION: System
```

```

: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/050.863
: FILING DATE: 30-MAR-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Silva, Robin M.
: REGISTRATION NUMBER: 38,304
: REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 949-8711
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2580 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
: US-09-050-863-2

Query Match      15.2%; Score 167.4; DB 3; Length 2580;
Best Local Similarity 53.0%; Pred. No. 2.2e-25;
Matches 382; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

Qy 7 ACCACCCAGAACCTCCACCTCTGACGCCATGTCAGCTCCTGCTGCTGCTGCTGCTG 66
Db 1397 ACCCGGCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338
Qy 67 TCTGACACAGAGCTGTGCTCAAGGTCTCGCGCCAGGAGAGCTGCTGCTGCTGCTGCTGCT 126
Db 1337 TCCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1278
Qy 127 CAGACCACTGCTGACGAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 186
Db 1277 TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1218
Qy 187 AGGCTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 246
Db 1217 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1158
Qy 247 AGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
Db 1157 TGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1101
Qy 307 ATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 366
Db 1100 CCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1041
Qy 367 ATTTCCAGTTGCTGTCAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 426
Db 1040 TCCTGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 981
Qy 427 AGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
Db 980 TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 921
Qy 487 ATCTCTAGCTGCTGTGCAAAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 546
Db 920 TCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 861
```



```
Qy 727 T 727
Db 726 T 726

RESULT 11
US-09-620-925-1/c
; Sequence 1, Application US/09620925
; Patent No. 6468986
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/620,925
; FILING DATE: 21-Jul-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/910,647
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujica, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9600 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1
Query Match 15.2%; Score 167.4; DB 4; Length 9600;
Rest Local Similarity 53.0%; Pred. No. 3.2e-25;
Matches 382; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

Qy 7 ACCACCCAGAACCTCCACCTCTGACGCCATGCTGCTGCTGGCTGTCTGTC 66
Db 1443 ACCCGGGCTCCACTGCTCTCTGACCGCGGCTCCACCTCTGCTGCGCCCTCTGTC 1384
Qy 67 TCTGACGAGAGCTGTGGTCAAGGTCTCGGCGCAGGAGAGCTGTCGCGCCAGCTGTGTC 126
Db 1383 TCCTGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1324
Qy 127 CAGACCACTCTGTCAGGACCACTGTCGCGCGCCAGCTGCTGATTCAGATTGCTGC 186
Db 1323 TCCTGCTCTGCGCCCTCTCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTCTGCTGTC 1264
Qy 187 AGGCCTTCTGCTGTATCTCCAGCTGTGTGCAAAACCCAGCTGCTGCTGACCACTGCTGC 246
Db 1263 TGCCCTCTGCGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1204
Qy 247 AGGACCACTGCTGCGCGCCAGCTGTGTGCAATTCAGATTGCTGTCGAGCCCTTCTGCTGCT 306
Db 1203 TGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1147

307 ATCTCCAGCTGCTGCAAAACCCAGCTGCTGAGGACCACTGCTGCGCGCCAGCTGCTGTC 366
1146 CCCTCTGCTGCTGCGCCCTCTGCTGCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCTGTC 1087
367 ATTTCCAGTTGCTGTCAGGCGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGTC 426
1086 TCCTGCGCCCTCTGCTGCTGCTGCGCCCTCTGCGCCCTCTGCTGCGCCCTCTGCTGCTGTC 1027
427 AGGACCACTGCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
1026 TCCTGCTGCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 967
487 ATCTCTAGCTGTTGCAAAACCCAGCTGCTGCGAGACCACTGCTGCGCGCCAGCTGCTGCTGCT 546
966 TCCTCTGCTGCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 907
547 ATCTCCAGCTGCTGACAGGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606
906 TGCTGCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
607 ATTTCTAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666
846 TGCGCCCTCTGCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
667 CCGACCACTGCTGCTGACAAACCCAGCTGCTGCGACCACTGCTGCGCGCCAGCTGCTGCTGCTGCT 726
786 TCCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
727 T 727
726 T 726

RESULT 12
US-07-884-811-15/c
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dregler, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 755.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
```

[illegible]

RESULT 13
US-07-885-971-15/c
: Sequence 15, Application US/07885971
: Patent No. 5328837
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
: TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 460 Point San Bruno Blvd
: CITY: South San Francisco

```

: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: patin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/885,971
: FILING DATE: 19920518
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dreger, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: 779
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-3216
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-07-885-971-15

Query Match 15.2%; Score 167.4; DB 1; Length 10596;
Best Local Similarity 53.0%; Pred. No. 3,3e-25;
Matches 382; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

Qy 7 ACCCACCAGAACCTCCACCGCTGTGACGCCATGTGTACAGCTCCTGCTGTGGCTTGTGTGTCG 66
Db 2977 ACCCGGCGCTCCACTGCTGTCTGCACCGCGGCGCTCCACCTCCTGCTGTGCGCCCTCCCTGC 2918
Qy 67 TCTCACCAGAGCTCTGTGTCAAGGTCTCGGCCAGGAGAGCTGCTGCGCGCCCAATGTGTGTCG 126
Db 2917 TCCGTGCGCCCTCCTCTCTGTCTGTGCGCCCTGTCTGCGCCCTGTCTGCTGCGCCCTGTGCGCC 2858
Qy 127 CAGACCACCTGCTGTCAGGACCACCTGCTGCGCGCCCGCCAGCTGCTGCATTTTCCAGTTGCTGTC 186
Db 2857 TCCTGCTCTCTGCGCCCTCTCTGCGCCCTGCTGTCTGCGCCCTCTCTGCGCCCTCTCTGCTGCTCC 2798
Qy 187 AGGCGTTCTCTGCTGTATCTTCAGAGTGTGCAAAACCCAGCTGCTGCTGTGACACCACTGCTGTC 246
Db 2797 TGCGCCCTGCTGCGCCCTCTCTGCTGTGCGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2738
Qy 247 AGGACACCTGCTGCGCGCCCGAGCTGTGCAATTTCCAGTTGCTGCGAGGCGCTTGTGCTGCTGT 306
Db 2737 TGCGCCCTGCTGCTGCTGCGC---CTCCTGCGCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2681
Qy 307 ATCTCCAGCTGCTGCAAAACCCAGCTGCTGTCAGGACCACCTGCTGCGCGCGCCCAATGTGTGTCG 366
Db 2680 CCCCTCTGCTGCTGCGCCCTCTGCTGCTGCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCTGCTGCT 2621
Qy 367 ATTTCCAGTTGCTGCGAGGCTTCTGCTGTATCTGATTTCCAGCTGCTGCTGCAAAACCAATGTGTGTC 426
Db 2620 TCTTGCGCCCTCTGCTGCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCT 2561
Qy 427 AGGACACCTGCTGCGCGCCCGAGCTGTGCAATTTCTAGTGTCTGCGAGGCGCTTGTGCTGCTGT 486
Db 2560 TCTGCTGCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2501
Qy 487 ATCTCTAGCTGTGCAAAACCCAGCTGCTGCGAGACCACCTGCTGCGCGCCCGCCAGTGTGTGT 546
Db 2500 TCTTCTGCTGCTGCGCCCTCTGCTGCTGCTGCGCCCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCT 2441
Qy 547 ATCTTCCAGCTGCTACAGGCGCCCGAGTGTGCGAGCGCCCTCTGCTGCTGCGCGCCCGGTGTGTGTC 606

```

[illegible]

Db 2440 TGCTCTGCCCTCTCTGCTGCCCTCTCTCTGCTGCCCTCTCTCTGCTGCCCTCTCC 2381
Qy 607 ATTTCTAGTGTCTCATCCAGCTGTGTGTCCAGCTGCCGTGCCCTTTTCAGCTGC 666
Db 2380 TGCCCTCTGCCCTCTCTGCTGCCCTCTCTCTGCTGCCCTCTCTCTGCTGCCCTCTCC 2321
Qy 667 CCGACCACTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCGGCAGTCTTGTGCTGC 726
Db 2320 TCTGCCCCCTCTGCTGCCCTCTCTGCTGCCCTCTCTGCTGCCCTCTCTGCTGCCCTCTG 2261
Qy 727 T 727
Db 2260 T 2260

RESULT 14
US-08-087-783A-15/c
; Sequence 15, Application US/08087783A
; Patent No. 5547856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087.783A
; FILING DATE: 13-Jul-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/885971
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,600
; REFERENCE/DOCKET NUMBER: P0755779P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5416
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-087-783A-15

Query Match 15.2%; Score 167.4; DB 1; Length 10596;
Best Local Similarity 53.0%; Pred. No. 3.3e-25;
Matches 382; Conservative 0; Mismatches 336; Indels 3; Gaps 1;

Qy 7 ACCACCCAGAACCTCCACCTCTGACGCCATGGTCAGCTGCTGTGGCTCTGCTGTC 66
Db 2977 ACCCGGCTCCACTGCTCTCTCGACCCCGGGCTCCACCTCTGCTGCTGCCCTCTG 2918
Qy 67 TGTGACCAAGAGTGTGGTCAAGGTCTGCGCCAGGAGAGCTGTCGCCGCCAGCTGCTGC 126
Db 2917 TCTGCCCCCTCTCTGCTGCTGCTGCCCTCTCTGCTGCCCTCTGCTGCCCTCTGCC 2858

Qy 127 CAGACCACCTGCTGCAAGSACCACTGCTGCTGCCGCCAGCTGCTGCAATTTCCAGTTGCTGC 186
Db 2857 TCCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCC 2798
Qy 187 AGGCTTTCCTGCTGTATCTATCTCAGCTGTGTGCAAAACCCAGCTGCTGCCGTGACACATGCTGTC 246
Db 2797 TGCCCTCTCTGCCCTCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCC 2738
Qy 247 AGGACCACCTGCTGCCGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306
Db 2737 TGCCCTCTGCTGCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCC 2681
Qy 307 ATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGCTGCCGCCAGCTGCTGCTGCCGCCAGCTGCTGCTGCC 366
Db 2680 CCGTCTGCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCC 2621
Qy 367 ATTTCCAGTTCCTGCAAGCCCTTCTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCTGCC 426
Db 2620 TCTGCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCC 2561
Qy 427 AGGACCACCTGCTGCCGCCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 486
Db 2560 TCCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCC 2501
Qy 487 ATCTCTAGCTGCTGCAAAACCCAGCTGCTGCTGCCGCCAGCTGCTGCTGCCGCCAGCTGCTGCTGCC 546
Db 2500 TCCTGCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCC 2441
Qy 547 ATCTGCACTGCTGCAAGCCGCCAGCTGCTGCTGCCGCCAGCTGCTGCTGCCGCCAGCTGCTGCTGCC 606
Db 2440 TGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCC 2381
Qy 607 ATTTCTAGTTCCTGCTGCTGCCAGCTGCTGCTGCTGCCAGCTGCTGCTGCCAGCTGCTGCTGCCAGCTGCTGCC 666
Db 2380 TGCCCTCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCTGCCCTCTGCTGCC 2321
Qy 667 CCGACCACTGCTGTAGAACCACTGCTTCCACCCCACTGCTGCTGCCGCCAGCTGCTGCTGCCGCCAGCTGCTGCTGCC 726
Db 2320 TCTGCCCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCC 2261
Qy 727 T 727
Db 2260 T 2260

RESULT 15
US-08-194-088B-15/c
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194.088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 03:06:50 ; Search time 325 Seconds
(without alignments)
7629.077 Million cell updates/sec

Title: US-09-874-062-2

Percent score: 1101
Sequence: 1 ctgggaaccaccaggaacc.....caaaaaaaaaaaaaaaaaa 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1101	100.0	1101	22	Human cDNA clone H
2	521.4	47.4	627	23	DNA encoding novel
3	278.6	25.3	1851	22	Human polynucleoti
4	278.6	25.3	1851	22	Human polynucleoti
5	252	22.9	3489	21	Kaposi's sarcoma-a
6	252	22.9	3489	22	Nucleotide sequenc
7	252	22.9	3489	22	Kaposi's sarcoma-a
8	252	22.9	3220	20	KSHV LUR DNA (nucl
9	252	22.9	137507	19	KSHV long unique c

ALIGNMENTS

RESULT 1

AAD06303

ID AAD06303 standard; cDNA: 1101 BP.

XX AAD06303;

AC AAD06303;

DT 10-AUG-2001 (first entry)

XX Human cDNA clone HHQC55 encoding FDCD-containing protein.

XX Human: four disulfide core domain; FDCD; immunosuppressive; cytostatic; antiarthritic; antirheumatic; antiproliferative; cardiast; vasotropic; cerebroprotective; neurotropic; dermatological; antifungal; virucide; fungicide; ophthalmological; dermatological; antineoplastic; antitumor; antiarteriosclerotic; gene therapy; skin disorder; congenital disorder; mole; freckle; haemangioma; integumentary tumour; basal cell carcinoma; keratosis; melanoma; atherosclerosis; urticaria; photosensitivity; eczema; autoimmune disorder; lupus erythematosus; scleroderma; keloid; stria; erythema; petechia; purpura; xanthelasma; autoimmune disease; rheumatoid arthritis; hyperproliferative disorder; cardiac arrest; cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia; neoplasm; angiogenesis; nervous system disorder; Alzheimer's disease; infection; ocular disorder; corneal infection; clone HHQC55; ds.

XX Homo sapiens.

OS DNA encoding novel

XX Human polynucleoti

XX Human polynucleoti

XX Kaposi's sarcoma-a

XX Nucleotide sequenc

XX Kaposi's sarcoma-a

XX KSHV LUR DNA (nucl

XX KSHV long unique c

Human hair keratin
Human polynucleoti
Human polynucleoti
Human foetal liver
Human brain expres
Human bone marrow
Probe #604 used t
Human genome-deriv
Oesophagus cancer
Epstein Barr virus
EBV tethering prot
Nucleotide sequenc
Epstein-Barr virus
Anti-sense strand
Vector pShuttle DN
Vector plasmid pCM
Nucleotide sequenc
Plasmid pCISERON f
Nucleotide sequenc
DNA clone pCEK Cl.
Nucleotide sequenc
Human foetal liver
Human brain expres
Human bone marrow
Probe #19224 used
Human genome-deriv
Epstein Barr Virus
Human foetal liver
Probe #614 for ge
Human brain expres
Human bone marrow
Probe #6199 for ge
Probe #7430 used t
Human genome-deriv
FLGA insert stabl

PN W0200140249-A1.
XX 07-JUN-2001.
XX 29-NOV-2000; 2000WO-US32462.
XX 01-DEC-1999; 99US-0168229.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Shi Y;
XX WPI: 2001-356153/37.
XX P-PSDB; AAE02058.
XX Four disulfide core domains (FDCD) containing polypeptide and its
PT polynucleotide are used to prevent, treat or ameliorate a medical
PT condition associated with FDCD e.g. skin disorders -
XX
PS Claim 1: Page 229; 235pp; English.
XX The present sequence is human cDNA clone HWHC55 encoding four
CC disulfide core domain (FDCD)-containing protein. FDCD sequences are
CC used to prevent, treat or ameliorate skin disorders such as congenital
CC disorders e.g. moles, freckles and haemangiomas as well as
CC intermediary tumours (keratosis, basal cell carcinoma, malignant
CC melanoma), injuries and inflammation of the skin, atherosclerosis,
CC urticaria, eczema, photosensitivity, autoimmune disorders (lupus
CC erythematosus, scleroderma), keloids, striae, erythema, petechiae,
CC purpura and xanthelasma. They are also used in the diagnosis and
CC treatment of diseases and disorders of hair and epithelium.
CC autoimmune diseases (e.g. rheumatoid arthritis), hyperproliferative
CC disorders (e.g. neoplasms of the breast or liver), cardiovascular
CC disorders (e.g. cardiac arrest), cerebrovascular disorders (e.g.
CC cerebral ischaemia, angioneurosis), nervous system disorders (e.g.
CC Alzheimer's disease), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infection). The FDCD-
CC containing proteins are used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. They are used in
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition. The antibodies to FDCD protein are used in alleviating
CC symptoms associated with the disorders and in diagnostic immunoassays
CC e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
CC The FDCD sequences are also used in gene therapy.
XX
SQ Sequence 1101 BP; 220 A; 343 C; 231 G; 307 T; 0 other;

Query Match 100.0%; Score 1101; DB 22; Length 1101;
Best Local Similarity 100.0%; Pred. No. 2.7e-194;
Matches 1101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGGGAACCCAGCCAGACCTCCACCTCTGACGCCATGCTCAGCTCTGCTGCTGCTCT 60
DB 1 CTGGGAACCCAGCCAGACCTCCACCTCTGACGCCATGCTCAGCTCTGCTGCTCT 60

QY 61 GTCTGCTGTGACAGAGCTGTGGTCAAGCTCTCGGCCAGGAGAGCTGTGCGGCCCCAGC 120
DB 61 GTCTGCTGTGACAGAGCTGTGGTCAAGCTCTCGGCCAGGAGAGCTGTGCGGCCCCAGC 120

QY 121 TGTGTCGACAGACCTGCTGTCAGGACCCTGCTGCGGCCCCAGCTGCTGCATTTCCAGT 180
DB 121 TGTGTCGACAGACCTGCTGTCAGGACCCTGCTGCGGCCCCAGCTGCTGCATTTCCAGT 180

QY 181 TGTGTCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGACACCC 240
DB 181 TGTGTCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGACACCC 240

QY 241 TGTGTCAGGACCACTGCTGCGGCCCCAGCTGCTGCATTTCCAGTGTGTCGACGGCTTCC 300
DB 241 TGTGTCAGGACCACTGCTGCGGCCCCAGCTGCTGCATTTCCAGTGTGTCGACGGCTTCC 300

QY 301 TGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGACGACACCTGCTGCGGCCAGC 360
DB 301 TGCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGACGACACCTGCTGCGGCCAGC 360

QY 361 TGCTGCATTTCCAGTTGCTGACAGGCTTCTGCTGCTATCTCCAGCTGCTGCAAAACCCAGC 420
DB 361 TGCTGCATTTCCAGTTGCTGACAGGCTTCTGCTGCTATCTCCAGCTGCTGCAAAACCCAGC 420

QY 421 TGCTGCAGGACCACTGCTGCGGCCCCAGCTGCTGCATTTCTAGTTGCTGCAGAGCTTTC 480
DB 421 TGCTGCAGGACCACTGCTGCGGCCCCAGCTGCTGCATTTCTAGTTGCTGCAGAGCTTTC 480

QY 481 TGCTGTATCTCTAGCTGTTGCAAAACCCAGCTGCTGCCAGACACCTGCTGCGGCCAGC 540
DB 481 TGCTGTATCTCTAGCTGTTGCAAAACCCAGCTGCTGCCAGACACCTGCTGCGGCCAGC 540

QY 541 TGCTGTATCTCCAGCTGCTACAGGCCCCAGCTGCTGCCAGGCCCTCTGCTGCGGCCAGCT 600
DB 541 TGCTGTATCTCCAGCTGCTACAGGCCCCAGCTGCTGCCAGGCCCTCTGCTGCGGCCAGCT 600

QY 601 TGCTGCATTTCTAGCTGTTGCTGTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTTTC 660
DB 601 TGCTGCATTTCTAGCTGTTGCTGTCATCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTTTC 660

QY 661 AGCTGCCGACCACTGCTGTAGAACCACTGCTTCCACCCCATCTGCTGCGGCCAGTTCT 720
DB 661 AGCTGCCGACCACTGCTGTAGAACCACTGCTTCCACCCCATCTGCTGCGGCCAGTTCT 720

QY 721 TGCTGCTGAGTGAACTGCTCTGGATTTGTCACCTTCTGCTCACAAGCTTTATTTCTTA 780
DB 721 TGCTGCTGAGTGAACTGCTCTGGATTTGTCACCTTCTGCTCACAAGCTTTATTTCTTA 780

QY 781 GCACAGAGTATCTATTTCAGAACCATGTTGAGCTTCTGATGCTGCGTGAACACAGAGCATG 840
DB 781 GCACAGAGTATCTATTTCAGAACCATGTTGAGCTTCTGATGCTGCGTGAACACAGAGCATG 840

QY 841 GACTGATTGGAAATATTTTATTAGTATGTTCTCTTTTATAGAAGTTTTTATTCTTA 900
DB 841 GACTGATTGGAAATATTTTATTAGTATGTTCTCTTTTATAGAAGTTTTTATTCTTA 900

QY 901 TTGAATCTGAATTTACAGTCAAAATTCACATCAGATGTTTGAATTTCTTTATTTCTTA 960
DB 901 TTGAATCTGAATTTACAGTCAAAATTCACATCAGATGTTTGAATTTCTTTATTTCTTA 960

QY 961 CAATATACATAAATCTTCAATGGTATCTTCTAGATGTTTCTTCTTAATGTTTCTG 1020
DB 961 CAATATACATAAATCTTCAATGGTATCTTCTAGATGTTTCTTCTTAATGTTTCTG 1020

QY 1021 GTATCAATTTTTCATGTTGAAATGTTTGTGATGTTCTCAATAAAGCTTCATAGTTTCAAAA 1080
DB 1021 GTATCAATTTTTCATGTTGAAATGTTTGTGATGTTCTCAATAAAGCTTCATAGTTTCAAAA 1080

QY 1081 GCAGAAAAAAGAAAAA 1101
DB 1081 GCAGAAAAAAGAAAAA 1101

RESULT 2
AAS93555
ID AAS93555 standard; cDNA; 627 BP.
XX
AC AAS93555;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29359.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN W0200175067-A2.

peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

XX Sequence 1851 BP; 379 A; 558 C; 574 G; 340 T; 0 other;

Query Match		25.3%	Score 278.6;	DB 22;	Length 1851;
Best Local Similarity		65.7%	Pred. No. 2.3e-42;		
Matches 480;		Conservative 0;	Mismatches 224;	Indels 27;	Gaps 4;
QY	17	AACCTCCACCCCTCTGACGCCATGGTCAGCTCCCTGCTGGTGTGCTGCTCTGTGACACAGA	76		
Db	933	AACCTCCACCCCTCTGACGCCATGGTTAACTCTGTTGGCTCTGCTCTGACACAGG	874		
QY	77	GCTGTGTTCAAGTCTCGCCAGGAGAGCTGCTGCGCCGCCAGCTGCTGCCAGACACCT	136		
Db	873	GCTGTGATCAAGGCCCTCTGCAAGAGACCTGCTGCGCCGCCAGCTGCTGCCAGACACCT	814		
QY	137	GCTGCAGGACCACTGCTGCGCCGCCAGCTGCTGTCATTTCCAGTTGCTGCAGGCCCTTCT	196		
Db	813	GTTCGTCGCCCACTGCTGTGTATCCAGCTGCTGCGCCGCCATCTCTCTCAGCTACCT	754		
QY	197	GCTGTATCTCCAGCTGCTGCAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	256		
Db	753	GCTGCCAGACCACTGCTGCTGCGCCGCCAGCTGCTGCGCCGCCAGCTGCTGCTGCTGCT	696		
QY	257	GCTGCGCGCCAGCTGCTGCATTTCCAGTTGCTGTCAGGCTTCTGCTGCTATCTCCAGCT	316		
Db	695	CTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	640		
QY	317	GCTGCAAAACCCAGCTGCTGCAGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	376		
Db	639	CTGCGCAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	583		
QY	377	GCTGCAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	436		
Db	582	CTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	523		
QY	437	GCTGCGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	496		
Db	522	GCTGCGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	463		
QY	497	GTTC-----CAAAACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	538		
Db	462	GGACTATCAACCATGAGCCAGTCACCATCCCATGATATGAAGAAGTCTGCTGCCGTCCTCA	403		
QY	539	GCTGCTGTATCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	598		
Db	402	GCTGCTGCCAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	343		
QY	599	CTTGCTGCATTTCTAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	658		
Db	342	CTGCTGTCAGACCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	283		
QY	659	TCAGCTGCCCGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	718		
Db	282	GCTGCTGCCAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	223		
QY	719	CTTGCTGCTGCA	729		
Db	222	GCTGCTTCA	212		

RESULT 4
AAK53119
ID AAK53119 standard; cDNA; 1851 BP.
XX

AAK53119;
06-NOV-2001 (first entry)
Human polynucleotide SEQ ID NO 2648.
Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
DR P-PSDB; AAM79986.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
useful in diagnosis and gene therapy -
Claim 1; Page 4903; 622pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and/or
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
(AAM80020) are omitted as the relevant pages from the sequence listing
were missing at the time of publication.
XX Sequence 1851 BP; 340 A; 574 C; 558 G; 379 T; 0 other;

Query Match		25.3%	Score 278.6;	DB 22;	Length 1851;
Best Local Similarity		65.7%	Pred. No. 2.3e-42;		
Matches 480;		Conservative 0;	Mismatches 224;	Indels 27;	Gaps 4;
QY	17	AACCTCCACCCCTCTGACGCCATGGTCAGCTCCCTGCTGGTGTGCTGCTCTGTGACACAGA	76		
Db	919	AACCTCCACCCCTCTGACGCCATGGTTAACTCTGTTGGCTCTGCTCTGACACAG	978		
QY	77	GCTGTGTTCAAGTCTCGCCAGGAGAGCTGCTGCGCCGCCAGCTGCTGCCAGACACCT	136		
Db	979	GCTGTGATCAAGGCCCTCTGCAAGAGACCTGCTGCGCCGCCAGCTGCTGCTGACACCT	1038		
QY	137	GCTGCAGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	196		

Db 1776 CCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTC 1717
 QY 462 TAGTTGCTGCGAGGCTTCTGCTGATCTAGCTGTTGCAAAACCCAGCTGCTGCACAG 521
 Db 1716 CCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTC 1657
 QY 522 CACCTGCTGCGCGCCCGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 581
 Db 1656 CCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTCCCGCTGCTGGCTC 1597
 QY 582 CTCCTGCTGCGCGCCCGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 641
 Db 1596 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1537
 QY 642 CAGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701
 Db 1536 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1477
 QY 702 CATCTGCTGCGCGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 761
 Db 1476 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1417
 QY 762 CTCT 765
 Db 1416 CTGT 1413

RESULT 6
 AAF82901/C
 ID AAF82901 standard; DNA; 3489 BP.
 XX AAF82901;
 XX 29-JUN-2001 (first entry)
 DT Nucleotide sequence of KSHV tethering protein, LANA.
 DE
 XX Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
 KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
 KW KSHV; latency-associated nuclear antigen; LANA; ds.
 XX
 OS Kaposi's sarcoma associated herpesvirus.
 XX
 FH Key
 FT 1..3489
 CDS /*tag= a
 XX
 XX WO200125484-A2.
 XX
 XX 12-APR-2001.
 XX
 XX 29-SEP-2000; 2000WO-US26908.
 XX
 XX 01-OCT-1999; 99US-0410399.
 XX
 XX (UNMI) UNIV MICHIGAN.
 XX
 XX Robertson ES, Cotter MA;
 FI
 FI WPI; 2001-281736/29.
 DR P-PSDB; AAB62331.
 DR
 XX
 PT A composition for use in gene therapy comprises an expression vector
 PT that includes a nucleic acid sequence encoding a nucleic acid binding
 PT protein .
 XX
 PS Disclosure; Fig 9A; 60pp; English.
 XX
 CC The invention provides a composition comprising nucleic acid, histone H1
 CC protein and expression vector operationally encoding a protein suitable
 CC for tethering the nucleic acid to the histone H1 protein, where the
 CC tethering protein is LANA. The composition is useful in aiding the
 CC retention of the viral DNA in the host cell. The viral vector encodes a

CC protein suitable for tethering DNA to Histone H1. Methods for screening
 CC for compounds which are agonistic or antagonistic for the tethering of
 CC viral proteins to histone H1 and DNA binding sites are useful for
 CC developing the method of viral transfer. The composition has applications
 CC to gene therapy, including the treatment of multiple sclerosis,
 CC Parkinson's disease, Huntington disease and diabetes. The present
 CC sequence represents the nucleotide sequence of the Kaposi's sarcoma
 CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear
 CC antigen (LANA), which acts as a tethering protein.
 XX
 SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;
 Query Match 22.9%; Score 252; DB 22; Length 3489;
 Best Local Similarity 59.3%; Pred. No. 2.2e-37;
 Matches 429; Conservative 0; Mismatches 295; Indels 0; Gaps 0;
 QY 42 CAGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 101
 Db 2136 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2077
 QY 102 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 161
 Db 2076 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2017
 QY 162 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 221
 Db 2016 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1957
 QY 222 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 281
 Db 1956 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1897
 QY 282 CAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 341
 Db 1896 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1837
 QY 342 CACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 401
 Db 1836 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1777
 QY 402 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 461
 Db 1776 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1717
 QY 462 TAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 521
 Db 1716 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1657
 QY 522 CACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 581
 Db 1656 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1597
 QY 582 CTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 641
 Db 1596 CTG 1537
 QY 642 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 701
 Db 1536 CTG 1477
 QY 702 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 761
 Db 1476 CTG 1417
 QY 762 CTCT 765
 Db 1416 CTGT 1413

RESULT 7
 ABA93487/c
 ID ABA93487 standard; DNA; 3489 BP.
 XX

[illegible]

355 GAACACCTCTTGCAGCCACCTGCTGTGGTCCAGCTGCTGCCAGCTTGTGTCACCC 296
444 CCCAGCTGTGATTTCTAGTTG-----CTGCAGGCTTCTGCTGTATCTCTAGCTG 497
295 AACATGCTGTCAACCAATTTGTAGATCCACCTGCTGCCAACCACTCTGTGTGACCATG 236
498 TTGCAAAACCCAGCTGTGTGCAGACCACTGCTGCGCGCCAGCTGTGTATCTCCAGCTG 557
235 CTGCAGCACACCTGTGTGCCAGCAACCTGTGTGGTCCAGCTGTGTAGCCAAACCTG 176
558 CTACAGGCCCCAGTGTGTGCCAGCCCTCTGCTGCGCGCCGCTGTGTATCTAGTTG 617
175 CAATGAGTCCAGCTATTGTCTGC---CTTGCTGCGCTCCACCTGCTGCGAGACCACTG 119
618 CTGTATCTCCAGCTGTGTGTGCCAGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 677
118 CTACAGGACCACTGTGTGCGCGCCAGCTGTGTGCTGAGTCTTGTGTGCTGCTGCTG 59
678 CTGTAGAACAACCTGCTTCCACCCCACTGCTGCGG 713
58 CTGCAGCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 23

RESULT 13
ABA57772
ID ABA57772 standard; DNA: 461 BP.
XX AC ABA57772;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #6077.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 1; SEQ ID NO 6077; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC foetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 461 BP; 74 A; 171 C; 107 G; 109 T; 0 other;
Query Match 18.0%; Score 198; DB 22; Length 461;
Best Local Similarity 71.2%; Pred. No. 1.3e-27;
Matches 306; Conservative 0; Mismatches 115; Indels 9; Gaps 3;
QY 2 TGGAAACCCACCCAGAACCTTCCACCTCTGACGCCATGGTCAGCTCCTGCTGTGGTCTG 61
DB 41 TGGAAACCTTACCTAGAACCTTCCACCTCTGACACCATGGTTAACTCTTGTGTGGTCTG 100
QY 62 TCTGTCTTACCAAGCTGTGTGTCAGGCTCTCGGCCAGGAGCTGCTGCGCGCCAGCT 121
DB 101 TCTGCTCTGACGAGGCTGTGATCAAGGCTCTGCCAAGAGACCTGCTGCCGCGCCAGCT 160
QY 122 GCTGCCAGACCACTGCTGCAAGACCACTGCTGCGCGCCAGCTGCTGCATTTTCAGTT 181
DB 161 GCTGCCAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 220
QY 182 GCTGCCAGGCTTCTGCTGTATCTCTCAGCTGCTGCAAAACCCAGCTGCTGCTGCAACCT 241
DB 221 GCTCTCAGACTACCTGCTGCCAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280
QY 242 GCTGCCAGACCACTGCTGCGCGCCAGCTGCTGCAATTTCCAGTTGCTGCTGCTGCTGCT 301
DB 281 GTTCTCAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 337
QY 302 GCTGTATCTCCAGCTGCTGCAAAACCCAGCTGCTGCCAGGACCACTGCTGCGCGCCAGCT 361
DB 338 GCTGTCCAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 394
QY 362 GCTGCAATTTCCAGCTGCTGCAAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
DB 395 GTTGTCCAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 451
QY 422 GCTGCCAGGAC 431
DB 452 GCTGTCCAGAC 461
RESULT 14
AAK05838
ID AAK05838 standard; DNA: 461 BP.
XX AC AAK05838;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 5829.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 1; SEQ ID NO 6077; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC foetal liver. The present sequence is a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

DR WPI: 2001-483446/52.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX
 XX Example 4; SEQ ID NO: 5829; 650pp + Sequence Listing; English.
 PS
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX
 XX Sequence 461 BP; 74 A; 171 C; 107 G; 109 T; 0 other;
 SQ
 Query Match 18.0%; Score 198; DB 22; Length 461;
 Best Local Similarity 71.2%; Pred. No. 1.3e-27;
 Matches 306; Conservative 0; Mismatches 115; Indels 9; Gaps 3;
 QY 2 TGGAAACCCAGACCTCCACCTCTGACGCCATGGTCAGCTCTCTGTGGCTCTG 61
 DB 41 TGGAAACCTACCTAGAACCTCCACCTCTGACACCATGGTTAACTCTGTGGCTCTG 100
 QY 62 TCTGCTGACACAGCTGTGTCAGGCTCTCGCCAGGAGCTGCTGCCGCCCCAGCT 121
 DB 101 TCTGCTGACACAGGCTGTGATCAAGGCTCTGCCAAGAGACCTGTGCCGCCCCAGCT 160
 QY 122 GCTGCCAGACCTGTGTCAGGACCTCTGCTGCCGCCAGCTGTGTCATTTCCAGTT 181
 DB 161 GCTGCCAGACCTGTGTCGCCAGCTGTGTTGTATCCAGCTGTGCCGCCATCCT 220
 QY 182 GCTGAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCAGCTGTGCTGACACCT 241
 DB 221 GCTCTCAGACTACCTGTGTCAGACCTCTGCTGCCGCCAGCTGTGCCGCCAGCT 280
 QY 242 GCTCAGGACCTGTGTCGCCGCCAGCTGCTGCTGCTGTCAGTTGCTGCCGCCAGCT 301
 DB 281 GTTGTGACAGACCA---CTGCCGCCAGCTGTGTTGTATCCAGCTGTGCCGCCAGCT 337
 QY 302 GCTGTATCTCCAGCTGTGCAAAACCAGCTGCTGCCAGGACCTGTGCTGCCGCCAGCT 361
 DB 338 GCTGTGACAGACCA---CTGCCGCCAGCTGTGTTGTATCCAGCTGTGCCGCCAGCT 394
 QY 362 GCTGATTTCCAGTTGCTGAGGCTTCTGCTGTATCTCCAGCTGTGCTGCAAAACCAGCT 421
 DB 395 GTTGTGACAGACCA---CTGCCGCCAGCTGTGTTGTATCCAGCTGTGCCGCCAGCT 451
 QY 422 GCTGAGGAC 431
 DB 452 GCTGTGACAG 461
 RESULT 15
 ID AAK31470 standard; DNA; 461 BP.
 XX
 AC AAK31470;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 6027.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX

PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-488900/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 XX Example 4; SEQ ID NO: 6027; 658pp + Sequence Listing; English.
 PS
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 XX Sequence 461 BP; 74 A; 171 C; 107 G; 109 T; 0 other;
 SQ
 Query Match 18.0%; Score 198; DB 22; Length 461;
 Best Local Similarity 71.2%; Pred. No. 1.3e-27;
 Matches 306; Conservative 0; Mismatches 115; Indels 9; Gaps 3;
 QY 2 TGGAAACCTACCTAGAACCTCCACCTCTGACGCCATGGTCAGCTCTCTGTGGCTCTG 61
 DB 41 TGGAAACCTACCTAGAACCTCCACCTCTGACACCATGGTTAACTCTGTGGCTCTG 100
 QY 62 TCTGCTGACACAGCTGTGTCAGGCTCTCGCCAGGAGCTGCTGCCGCCCCAGCT 121
 DB 101 TCTGCTGACACAGGCTGTGATCAAGGCTCTGCCAAGAGACCTGTGCCGCCCCAGCT 160
 QY 122 GCTGCCAGACCTGTGTCAGGACCTCTGCTGCCGCCAGCTGTGTCATTTCCAGTT 181
 DB 161 GCTGCCAGACCTGTGTCGCCAGCTGTGTTGTATCCAGCTGTGCCGCCATCCT 220
 QY 182 GCTCAGGCTTCTGCTGTATCTCCAGCTGCTGCAAAACCAGCTGTGCTGCAAAACCAGCT 241
 DB 221 GCTCTCAGACTACCTGTGTCAGACCTCTGCTGCCGCCAGCTGTGCCGCCAGCT 280
 QY 242 GCTCAGGACCTGTGTCGCCGCCAGCTGCTGCTGCTGTCAGTTGCTGCCGCCAGCT 301
 DB 281 GTTGTGACAGACCA---CTGCCGCCAGCTGTGTTGTATCCAGCTGTGCCGCCAGCT 337
 QY 302 GCTGTATCTCCAGCTGTGCAAAACCAGCTGCTGCCAGGACCTGTGCTGCCGCCAGCT 361
 DB 338 GCTGTGACAGACCA---CTGCCGCCAGCTGTGTTGTATCCAGCTGTGCCGCCAGCT 394
 QY 362 GCTGATTTCCAGTTGCTGAGGCTTCTGCTGTATCTCCAGCTGTGCTGCAAAACCAGCT 421
 DB 395 GTTGTGACAGACCA---CTGCCGCCAGCTGTGTTGTATCCAGCTGTGCCGCCAGCT 451
 QY 422 GCTGAGGAC 431
 DB 452 GCTGTGACAG 461
 Search completed: June 18, 2003, 10:20:21
 Job time : 337 secs


```
QY 161 QT-----TCCR-----PSCCISCYRPQCC---OPSCCRPACCISCCCHPSOCVSSCRCPFS 210
DB 3122 ATATCTCCACTGCTACCTTGCTGCGCCAGAGACC--ACCTGCGCCAAACCAAC-CACTC 3178
QY 211 P-----TTCRTTCFHPICCGSSCC 230
DB 3179 AGGTAACGCCACTAATCAGCAGGGGCC 3206

RESULT 4
US-10-123-155-99
; Sequence 99, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Sherwood, Steven
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 99
; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-99

Query Match 32.3%; Score 479.5; DB 9; Length 1904;
Best Local Similarity 30.8%; Pred. No. 1.le-23;
Matches 94; Conservative 22; Mismatches 90; Indels 99; Gaps 19;

QY 3 SSCGSGVSDQSCG-----QCLGQESCRPSCCVTTCC-----RTTCGRP 42
DB 1397 TTTCACCTTGCGCGTAGGCTGG--CACTCCCTTCCCGCAGCTTTAATNACTCCTG 1453
QY 43 SCC---ISSCCRPSCCISCCKPSCLTTCGRTTCRPSCCISSC-CRPSCCISCCCKPS 98
DB 1454 GCCTGGCACCTCAACCCACCGCTACTTTCCCATCC---CCAGCGCTGTGCTCTTCA 1510
QY 99 CCRPTCCRPSCISS-----CCRPS-----CCISSCKPSCCRPTCCRPS 138
DB 1511 CCATACCCCGCTAGAGACTGTAAAGCGCTAAAGCCTCGGCTGTCC---TCC-CACCATT 1567
QY 139 C--CISSCCRPS--CCISSCKPSCCQT-----TCCRPSCCIS--CY 175
DB 1568 CTGCTGCGCATGCTGCTGCC---CCTTTCTCCAACCCCTATTAGGTACCGGAACA 1624
QY 176 RPQCC-----QPSCCRPACCISCCCHPSOCVSSCRCPFSCTTTCRTTCFHPITCC 225
DB 1625 GAACCCCTGGGCTGAGGCGCTGGCCCTGCC---CCCGGCC---CTGCCCTGTC----CC 1674
QY 226 GSSCC 230
DB 1675 GCCCC 1679
```

```
RESULT 5
US-10-184-644-281
; Sequence 281, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 281
; LENGTH: 1732
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-281

Query Match 32.3%; Score 479; DB 9; Length 1732;
Best Local Similarity 33.2%; Pred. No. 1.le-23;
Matches 97; Conservative 27; Mismatches 72; Indels 96; Gaps 21;

QY 5 CCGSGVSDQSCGQGLGQESCRPSCCQTTC-----CRTTCGRP----- 42
DB 2 CCACGCG--TCCGCG-----CCTCCCTCTTCGTCGACCTTCCTTCGTCCTCCATCTCT 55
QY 43 ---SCCISCCRPSC-----CISSCKPSCCLTTCGRTTC--CRPSCCIS--SCCRPS-C 90
DB 56 CCCTCCTTTCCCGGCTCTCTTTCC--ACCTTCTCTCTTCCACCTTAGAGC--TCC 111
QY 91 ISSCKPSCRTTCRPSCCISSCCRPSC-----ISSCKPSCRT--TCCRPSCISSC 144
DB 112 CTTCCTGCGC--TCTTTCTGCGCC--ACCGCTGCTCTCGCCCTTCTCCGACCCCT 167
QY 145 CRPSC--CISSCKRPSCCQ-----TT-----CCTRPSCISSCYRQVC 179
DB 168 CTAGCAGCAGAGC--TCTGGGCTGTGTGGTTCATCTGTGCGCCTCTGTGCTTCTGTC 225
QY 180 CUPSCKRP-ACCISCCCHPSOCVSSCRCPFSCTTTCRTTCFHPICCGSSC 240
DB 226 CTTTCTGCTGCTCTTCG--TCCGAG-----TCCGCTC-----CCGAGC 263

RESULT 6
US-10-184-634-281
; Sequence 281, Application US/10184634
; Publication No. US2003006864A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
```

```

: CURRENT APPLICATION NUMBER: US/10/184,634
: CURRENT FILING DATE: 2002-06-28
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 281
: LENGTH: 1732
: TYPE: DNA
: ORGANISM: Homo Sapien
US-10-184-634-281

Query Match          32.3%; Score 479; DB 9; Length 1732;
Best Local Similarity 33.2%; pred. No.1.le-23;
Matches 97; Conservative 27; Mismatches 72; Indels 96; Gaps 21

QY 5 CCGSVCSDDCGGGLGQESCRPSCCQTTC-----CRTTCRRP----- 42
Db 2 CCAGCG-TCGCG- ----CTCTCCCTCTGCTGGACCTTCCTTGCTCTCCATCTCT 55

QY 43 ---SCCISSCCRPSC-----CISCCCKPSCGLTTCCRTTC--CRPSCCIS--SCCRPSCC 90
Db 56 CCCTCCCTTCCCGCGGTCTCTTTCC--ACCTTCTCTCTTCCACCTTAGACC--TCC 111

QY 91 ISSCCKPSCRTTCRRPSCCISSCCRPSC-----ISSCCKPSCRT--TCRRPSCCISSC 144
Db 112 CTTCCTGCC--TCCTTTCTCTGCC--ACCGCTGCTCTCTGGCCCTTCTCGGACCCGCT 167

QY 145 CRPSC--CISCCCKPSCCO-----TT-----TCRRPSCCISSCYRQC 179
Db 168 CTACGACGAGACC--TCCTGGGTCTGTGGGTGATCTGTGCCCTGTGCTCGGTGTC 225

QY 180 COPSCRRP-ACCISCCHPSCCVSCRCRPFSCPTTCCMTCTCFHPICGSSCC 230
Db 226 CTTTTCGTCTCCCTTCC--TCCCGAC-----TCCGCTC---CGGGACC 263

```

RESULT 7
 US-09-864-761-39564
 : Sequence 39564, Application US/09864761
 : Patent No. US20020048763A1
 : GENERAL INFORMATION:
 : APPLICANT: Penn, Sharon G.
 : APPLICANT: Rank, David R.
 : APPLICANT: Hankel, David K.
 : APPLICANT: Chen, Wensheng
 : TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 : TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 : FILE REFERENCE: Aemica-X-1
 : CURRENT APPLICATION NUMBER: US/09/864,761
 : CURRENT FILING DATE: 2001-05-23
 : PRIOR APPLICATION NUMBER: US 60/180,312
 : PRIOR FILING DATE: 2000-02-04
 : PRIOR APPLICATION NUMBER: US 60/207,456
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR APPLICATION NUMBER: US 09/632,366
 : PRIOR FILING DATE: 2000-08-03
 : PRIOR APPLICATION NUMBER: GB 24263.6
 : PRIOR FILING DATE: 2000-10-04
 : PRIOR APPLICATION NUMBER: US 60/236,359
 : PRIOR FILING DATE: 2000-09-27
 : PRIOR APPLICATION NUMBER: PCT/US01/00666
 : PRIOR FILING DATE: 2001-01-30
 : PRIOR APPLICATION NUMBER: PCT/US01/00667
 : PRIOR FILING DATE: 2001-01-30
 : PRIOR APPLICATION NUMBER: PCT/US01/00664
 : PRIOR FILING DATE: 2001-01-30
 : PRIOR APPLICATION NUMBER: PCT/US01/00669
 : PRIOR FILING DATE: 2001-01-30
 : PRIOR APPLICATION NUMBER: PCT/US01/00665
 : PRIOR FILING DATE: 2001-01-30
 : PRIOR APPLICATION NUMBER: PCT/US01/00668
 : PRIOR FILING DATE: 2001-01-30
 : PRIOR APPLICATION NUMBER: PCT/US01/00663
 : PRIOR FILING DATE: 2001-01-30

```

> PRIOR APPLICATION NUMBER: PCT/US01/00662
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00661
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00670
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: US 60/234,687
> PRIOR FILING DATE: 2000-09-21
> PRIOR APPLICATION NUMBER: US 09/608,408
> PRIOR FILING DATE: 2000-06-30
> PRIOR APPLICATION NUMBER: US 09/774,203
> PRIOR FILING DATE: 2001-01-29
> NUMBER OF SEQ ID NOS: 49117
> SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
> SEQ ID NO 39564
> LENGTH: 134
> TYPE: PRT
> ORGANISM: Homo sapiens
> FEATURE:
> OTHER INFORMATION: MAP TO AC006070.1
> OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
> OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
> OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
> OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5
> OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
> OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
> OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
> OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
> OTHER INFORMATION: SWISSPROT HIT: P02438, EVALUE 7.40e-02
US-09-864-761-39564

Query Match          31.4%; Score 466.5; DB 10; Length 134;
Best Local Similarity 46.0%; Pred. No. 1.5e-23;
Matches 80; Conservative 19; Mismatches 34; Indels 41; Gaps 6;

Qy      25 CRPSCCGTTTCRTTTCRPSCCISSCCKPSCCLTTCCKTTCRPSGCCISSC 84
Dg      1   ||||| ||||| :|
       1 CSP-CCOPTCCRTCKMKT-----TWITCSSTPCCGPSCGVSSC 38

Qy      85 CRPSCCISSCKPSCRTTTCRPSCCISSCCRPSCCGTTCCKRPSGCCISSC 144
Dg      1   ||| :||| :||||| :|
       39 CUP-CGIPTCCQNVICHTCCQPT-----CVTSCCPSCSGTPCCGPICGSSC 86

Qy      145 CRPSCCISSCKPSCGTTTCRPSCC-ISSCYRPGCCQPSCCRPACTISGCHP 197
Dg      1   : ||||| ||| :| :| :| :| :| :| :| :| :| :| :| :| :|
       87 CGUTSUGSCGQPSSCAPIVCRKTCYHPTSVYLPGCLNQSCG-----SSCGVP 144

RESULT 8
US-09-864-761-43621
> Sequence 43621, Application US/09864761
> Patent No. US20020048763A1
> GENERAL INFORMATION:
> APPLICANT: Penn, Sharron G.
> APPLICANT: Rank, David K.
> APPLICANT: Hanzel, David K.
> APPLICANT: Chen, Weisheng
> TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
> TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
> FILE REFERENCE: Acomica-x-1
> CURRENT APPLICATION NUMBER: US/09/864,761
> CURRENT FILING DATE: 2001-05-23
> PRIOR APPLICATION NUMBER: US 60/180,312
> PRIOR FILING DATE: 2000-02-04
> PRIOR APPLICATION NUMBER: US 60/207,456
> PRIOR FILING DATE: 2000-05-26
> PRIOR APPLICATION NUMBER: US 09/632,366
> PRIOR FILING DATE: 2000-08-03
> PRIOR APPLICATION NUMBER: GB 24263.6
> PRIOR FILING DATE: 2000-10-04
> PRIOR APPLICATION NUMBER: US 60/236,359
> PRIOR FILING DATE: 2000-09-27
> PRIOR APPLICATION NUMBER: PCT/US01/00666
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43621
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006070.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EST_HUMAN HIT: BE250684.1, EVALUATE 2.00e-12
; OTHER INFORMATION: SWISSPROT HIT: O75690, EVALUATE 3.00e-10
US-09-864-761-43621

Query Match          30.2%; Score 448; DB 10; Length 102;
Best Local Similarity 51.3%; Pred. No. 2e-22;
Matches 77; Conservative 11; Mismatches 14; Indels 48; Gaps 5;

QY 25 CRPSCCQTTCCRTTCRPSCCISCCCRPSCCISCCCRPSCCISCCCRPSCCISCCCRPSCCISCC 84
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CRPSCQTTCCQTTCCRP-----SCCHPVCCQTT----- 29

QY 85 CRPSCCISCCCKPSCRTTCRPSCCISCCCRPSCCISCCCRPSCCISCCCRPSCCISCCCRPSCC 144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30 CRPSCGVSSCCRPCCQTT-----CHPSCGMSSCCRPCCQTT-CRPSCGVSSC 77
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 145 CRPSCCISCCCKPSCCQTTTCRPSCCISCC 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78 CRP-----LCCQTTCCRTTCRPSCCGSSC 102

RESULT 9
US-10-184-644-283
; Sequence 283, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 283
; LENGTH: 2294
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-283

Query Match          29.9%; Score 443; DB 9; Length 2294;
Best Local Similarity 31.5%; Pred. No. 2.9e-21;
Matches 96; Conservative 20; Mismatches 85; Indels 104; Gaps 22;

QY 4 SCGGSVCSQSGAGAGQESCCRP--SCCQTTCCRT-----TCCRSQCLSSC 49
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 333 ACCGGGCC-CCCGCGAGCGACCGTCCACGACCCCTGGCTGGCAGCTTCTCCAGAGCAATC 391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 50 CRPSQC--ISSCKPSCCLTFT-----CCR-TTCCRPSCCLSSCKRPSC 89
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 C---CGGAGACCAACCCCTCTTTGGGCGACTGCTGGACCCCTCTTC--AGC--AGCTTTC 444
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 90 -----CLSSCKPSCRTTCRPSCCISCCCRPSCC-----ISSCKP 127
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 445 AGCGGCGCTCGGCGCCCTCGCGACCAACCC--TCCGGGGCGGAGACGACCTTCAGAGC-- 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 SCCKTTC-----CRPSCLSSCKPSCCLSSCKPSCCQTTTC-----CRPSCLSSC 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 501 ACTCTCAGAGCGGAGACAGACCGCGCCGACCC-ACCCTTCGAGCAGCACTGAGAGC 559
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 YRPCCQCPSCRPACCLIS-----SCCHPSCCVSSCRCPSC-PTTCRTTCRTTCRTTC 225
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 560 ---GCC-GACCAACCCCTGTAGCGAGCAACCCCTACC-GCGCGCCAGCACTTC--CC 606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 226 GSSCC 230
    |||||
Db 607 GGACC 611

RESULT 10
US-10-184-634-283
; Sequence 283, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 283
; LENGTH: 2294
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-283
```

```
Query Match          29.9%; Score 443; DB 9; Length 2294;
Best Local Similarity 31.5%; Pred. No. 2.5e-21;
Matches 96; Conservative 20; Mismatches 85; Indels 104; Gaps 22;

QY 4 SCGSGVSDSGGGLQGESCRP--SCCOTTCRT-----TCCRPSCCISSC 49
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 ACCGGGCC-CCCGGGCCACCGTCGACGACCGCTGGTGGGACTTCTCCAGCCCGAGTC 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 50 CRPSCC--ISSCKPSCCLTT-----CCR--TTCRPSCCISSCRPSC 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 392 C---CCGGAGACACCCCTTTTGGGCGACTGTGGACCCCTCTTC--ACC--ACCTTTC 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 90 -----CTSSCKPSCCRTTCCRPSCCISSCRPSCC-----ISSCKP 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 AGCGCGCTCGGCCCTCGCCGACACCC--TCCGGCGGGGAGCGACATTCGACC-- 500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 128 SCRTTTC-----CRPSCCISSCRPSCCISSCRPSCCOTTC-----CRPSCCISSC 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 ACCTCTCAGCGCGGACGACGACCGCGCGGACC-ACCCTTTCGAGGACCACTGGCCGGC 559
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 175 YRPOCCOPSCRPACIS-----SCCHPSCVSCRCRPFSC-PTTCRTTTCFHPIC 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 560 ---GCC-GACCACCCCTGTAGCGACCACTAGC-GGCGCCACGACTCC-----CC 606
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 226 GSSCC 230
   ||| |||
Db 607 GGACC 611

RESULT 11
US-10-184-644-135
; Sequence 135, Application US/10184644
; Publication No. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 135
; LENGTH: 2651
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-135

Query Match          29.6%; Score 439; DB 9; Length 2651;
Best Local Similarity 32.4%; Pred. No. 5e-21;
Matches 91; Conservative 12; Mismatches 108; Indels 70; Gaps 18;

QY 5 CCGSVCSDDSG-OGLGQE-----SCCRPSCCOTTCRTTCCRPSCCISSCRPSCC 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 CCAAGCAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 56 ISSCKPSCCLTTCCRTTCCRPSCC--ISSCRPSCCISSCCKPSCCCTTC--CRPSC- 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 -TGCC--ACCTTTC-CTCCTTCCCGCGCTCCCGCGCTCCGC--GGCCAGTCAGTTCGCG 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 111 -----ISSCKRPSCCISSCC-----KPSCCRTTCCRPSCCISSCCKPSCCIS- 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 GGTTCGCTGCGCGCGGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 153 --SCCKPSCCOTTCRTTCCRPSCCISSCYRPOCCQPSCCRPAACCISSCKHPSCCVSS- 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 GCGCC--GCC--TCCAGCGCGCTCTCTTC--CCTGGCGCGGC--GCCCTGGTACGTGG 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 207 PFSCPTT-----CCRTTTCFHPIC-----CGSSCC 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 GACCGTTTGGCTGACGCGGAGCGCGAGCTCTACTTTTTCGCCCC 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-10-184-634-135
; Sequence 135, Application US/10184634
; Publication No. US2003006868A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 135
; LENGTH: 2651
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-135

Query Match          29.6%; Score 439; DB 9; Length 2651;
Best Local Similarity 32.4%; Pred. No. 5e-21;
Matches 91; Conservative 12; Mismatches 108; Indels 70; Gaps 18;

QY 5 CCGSVCSDDSG-OGLGQE-----SCCRPSCCOTTCRTTCCRPSCCISSCRPSCC 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 CCAAGCAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 56 ISSCKPSCCLTTCCRTTCCRPSCC--ISSCRPSCCISSCCKPSCCCTTC--CRPSC- 110
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 -TGCC--ACCTTTC-CTCCTTCCCGCGCTCCCGCGCTCCGC--GGCCAGTCAGTTCGCG 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 111 -----ISSCKRPSCCISSCC-----KPSCCRTTCCRPSCCISSCCKPSCCIS- 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 GGTTCGCTGCGCGCGGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 153 --SCCKPSCCOTTCRTTCCRPSCCISSCYRPOCCQPSCCRPAACCISSCKHPSCCVSS- 206
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 GCGCC--GCC--TCCAGCGCGCTCTCTTC--CCTGGCGCGGC--GCCCTGGTACGTGG 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 207 PFSCPTT-----CCRTTTCFHPIC-----CGSSCC 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 GACCGTTTGGCTGACGCGGAGCGCGAGCTCTACTTTTTCGCCCC 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-10-123-155-543
; Sequence 543, Application US/10123155
; Publication No. US2003006879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeBoeye, Laura
```

```
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 543
; LENGTH: 3721
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-543

Query Match      29.5%   Score 438.5; DB 9; Length 3721;
Best Local Similarity 32.4%; Pred. No. 6.5e-21;
Matches 89; Conservative 12; Mismatches 111; Indels 63; Gaps 17;

QY 5 CCGSVCSQDQCGGGLGQESCCRPSC-----CCQTTTCRTTCCRPSCGCGISCCRPSCCIS 58
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 2147 CCGAGCAG-ACG-CCGTCGCGGAGAGCGCGCTGCCCACTCCGACCCGCGAC---CC--- 2198

QY 59 CCRPSCCLT-----TCRTTCCR-----PSCCISCCRP--SCCISCCRPSCRTT 103
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 2199 CCAGCGCTGGCGCCCGCGCGCTGGGACACGCGGCGCGCGCTGCTCCCGCGC---TCGCGT 2256

QY 104 CCRPSC-CISSCCRP-----SCCISCCRPSCRTTCCRPSC--CISCCRPSCCI- 151
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 2257 TCATCTCTCTCTGCTGCTGCTGCGCGCGCGCGCGCGCGC---CCGAGCAGCCCGCGCGCTG 2314

QY 152 ---SSCCRPSCCQTTCCRPSCCISCCYRPQCCQRPACCTSSCCHPSCCVSSCRCP- 207
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 2315 GGGAGCGGACCGCGGCGCGCGCTCTATGC--TGCCCGCGCGCGCGCGCGCTCCCGCGG 2372

QY 208 -FSCPTTCRTT-----CFHPICGSSCC 230
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 2373 CGACTTCCGCTCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2407

RESULT 14
US-10-123-155-337
; Sequence 337, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
```

```
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 337
; LENGTH: 1813
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-337

Query Match      29.5%   Score 437.5; DB 9; Length 1813;
Best Local Similarity 29.9%; Pred. No. 4.9e-21;
Matches 93; Conservative 15; Mismatches 108; Indels 95; Gaps 18;

QY 4 SCGSVCSQDQCGGGLGQESCCRPSCCQTT-----GLGQESCCRPSCCQTT-----CCRTTCC 40
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 940 ACCG-CGCGGCGAGAGCGCGCTGGAGAGCGC-AGCAATGATATCAAGGAGGATGCGCATGC 997

QY 41 RPSCCISCCRPSCCIS-----SCCKPS-----SCCKPS-----SCCKPS----- 69
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 998 TCCCCGAGCGCTGCCCTGGCCCAAGAGCTCAGACACAATCTCCCAAGAAATGGAGAGCTTTC 1057

QY 70 CR-----TTCCKPSC---CISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 110
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1058 CTCTGTCTACCTCCGACGAGCGCTCC--GCGCACCCCATGGCCCTCCGACGCTGGTGTGAT 1116

QY 111 ISSCCRPSCCISCCRPSCCCTTCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 170
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1117 TGACC--CCGAGCGCGAGCTCTCTCCAGCGAGCGCTCCCTCACCAGAGCTGCGAGAGAG 1174

QY 171 ISSCRPQCCQPSCCR-PAGC--ISSCHPSCCVSSCR-----PFSCTTCCTTTC 219
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1175 AGATGGCGCGCGCGCTCAACCAATATCC---CCCATCCCTGGTGGGTTCCTTTC--TTC 1228

QY 220 FHPICGSSCC 230
   |||: | : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1229 TGGCTTGAGTC 1239

RESULT 15
US-10-123-155-381
; Sequence 381, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
```



```

:   LENGTH: 2236
:   TYPE: DNA
:   ORGANISM: Homo Sapien
US-10-123-155-381

Query Match      29.4%; Score 437; DB 9; Length 2236;
Best Local Similarity 32.2%; Pred. No. 6e-21;
Matches 89; Conservative 9; Mismatches 110; Indels 68; Gaps 16;

QY  5  CCGSVCDQSCGGLGQESCRPSCCOTTCCRTTCCRPSCCISCCRPSCCISCCCKPSC 64
Db   ||| |::||| || || || || || || || || || || || || || || ||
QY  40 CCGGCTGGCGCCCGCGCC--GCCGCGCGCCACGCCCAACC---CGGCGCGCGC 94
Db   ||| |::||| || || || || || || || || || || || || || || ||
QY  65 CLTTCRTTCCRPSCCISCCRPSC-----ISSCKPSC-----CRTTCCRP 107
Db   | || || || || || || || || || || || || || || || || || ||
QY  95 C--CCCTAGCC--CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 149
Db   | || || || || || || || || || || || || || || || || || ||
QY  108 SCCISSCRPSCCISCCCKPSCRTTCCRPSCCISCCRPSCCISCC-----KPSCC 160
Db   | || || || || || || || || || || || || || || || || || ||
QY  150 CCGCGCGGAGGCGCGCC--CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGGAACC 206
Db   |||| |::||| || || || || || || || || || || || || || || ||
QY  161 -----QTTCCRPSCCISCYRPQCCQPSCHPA--CC-----ISSCHPSCCVS 202
Db   |||| |::||| || || || || || || || || || || || || || || ||
QY  207 GGGCGGATTCCTCGCGGTC-AAACC--ACCTGATCCCATAAACATTCATCC--TCCCG 261
Db   |||| |::||| || || || || || || || || || || || || || || ||
QY  203 SC-RCPFSCPTT-----CCRTTCFHPICCGSSCC 230
Db   | || || || || || || || || || || || || || || || || || ||
QY  262 GCGGCGCGGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 297
Db   | || || || || || || || || || || || || || || || || || ||
```

Search completed: June 11, 2003, 11:14:44
Job time : 46 secs


```

QY 101 RTCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 160
Db 94 QCCYKRP-CCCSGGCGSCCQSCCKPCCSCCKP-CCSCGGSCCQSCCKPCCS 151
QY 161 QTCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCC 200
Db 152 QSSCCRPCCSSGC-GSSCCSCCKPCCSCCVPICC 190

RESULT 2
KR2D_SHEEP
ID KR2D_SHEEP STANDARD: PRT: 181 AA.
AC P08131.
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Keratin, high-sulfur matrix protein, B2D.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8329218; PubMed=6193483;
RA Powell B.C., Sleight M.J., Ward K.A., Rogers G.E.;
RT "Mammalian keratin gene families: organization of genes coding for
RT the B2 high-sulfur proteins of sheep wool.";
RL Nucleic Acids Res. 11:5327-5346(1983).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 KDA, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kda).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X01610; CAA35759.1; -.
DR PIR: S07911; KRSHHD.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin; Repeat.
FT INIT_MET 0
FT DOMAIN 26 85 6 x 10 AA TANDEM REPEATS.
FT REPEAT 26 35 1.
FT REPEAT 36 45 2.
FT REPEAT 46 55 3.
FT REPEAT 56 65 4.
FT REPEAT 66 75 5.
FT REPEAT 76 85 6.
SQ SEQUENCE 181 AA; 18679 MW; D3E4874E21757B12 CRC64;

Query Match 29.2%; Score 434; DB 1; Length 181;
Best Local Similarity 37.1%; Pred. No. 5.8e-29;
Matches 76; Conservative 27; Mismatches 54; Indels 48; Gaps 7;

QY 4 SCCG-SVCSQCGGLGQESCRCQTCRTCCRPSCCISCCRPSCCISCCRPSCCISCCCKP 62
Db 1 ACCTSFGCPPTSTG---GTGSGNFCOPTCCQTSCTCCOPTSIQTSQCCQP 56
QY 63 SCCLTTCRTTCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCCCKP 104
Db 57 -----TSIQTSQCCOPTSIQTSQCCOPTCLQTSGETGCGIGSGYGVGVSSGAVSSRTKW 111
QY 105 CRPSCCISCCRPSCCISCCCKPSCC-----RTTCRPSCCISCCRPSCCISCCCKPSC 159

```

```

Db 112 CRPDCRVECTSLPPCCVWSCTSPSCQLYVAQASCRPSYCGSCCRPAC-----C 162
QY 160 COTTCRRPSCCISCCYRPOCCPSC 184
Db 163 CQPTCIEPVC-----EPTCCPTC 181

RESULT 3
KRUA_HUMAN
ID KRUA_HUMAN STANDARD: PRT: 169 AA.
AC P26371.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, ultra high-sulfur matrix protein A (UHS keratin A) (UHS
DE KerA).
GN KRNI OR UHSK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91115951; PubMed=1703541;
RA McKinnon P.J., Powell B.C., Rogers G.E.;
RT "Structure and expression of genes for a class of cysteine-rich
RT proteins of the cuticle layers of differentiating wool and hair
RT follicles.";
RL J. Cell Biol. 111:2587-2600(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99148005; PubMed=10023043;
RA Perez C., Aurioi J., Gerst C., Bernard B.A., Egly J.-M.;
RT "Genomic organization and promoter characterization of two human UHS
RT keratin genes.";
RL Gene 227:137-148(1999).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 KDA, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kda).
CC -----
CC -!- TISSUE SPECIFICITY: CUTICLE LAYERS OF DIFFERENTIATING
CC HAIR FOLLICLES.
CC -!- DEVELOPMENTAL STAGE: AT A LATE STAGE OF FIBER DIFFERENTIATION.
CC -!- DOMAIN: MAINLY COMPOSED OF CYS-RICH (CR), GLY-RICH (GR) AND SER-
CC RICH (SR) REPEATS.
CC -!- SIMILARITY: BELONGS TO THE UHS KERATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55293; CAA39005.1; -.
DR EMBL: AJ006693; CAA07189.1; -.
DR HSSP: P01064; IPI2.
DR Genew: HGNC:6409; KRNI.
DR MIM: 148021; -.
KW Keratin; Repeat; Multigene family.
SQ SEQUENCE 169 AA; 16276 MW; 219B14FEB49DAAB CRC64;

Query Match 28.4%; Score 420.5; DB 1; Length 169;
Best Local Similarity 46.4%; Pred. No. 6.8e-28;
Matches 81; Conservative 13; Mismatches 56; Indels 25; Gaps 8;

QY 3 SSCGSGVCS--DQSGVGLGQESC-----CRPSCCQTCRTCCRPSCCISCCCKPSC 54

```



```
AC P02442;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Keratin, high-sulfur matrix protein, IIA3 (M2.6 protein).
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.
RC STRAIN=South African angora;
RA Joubert F.J.;
RT "Studies on the high-sulphur proteins of reduced mohair. The
RT isolation and the amino acid sequence of protein SCMB-M2.6.";
RL J. S. Afr. Chem. Inst. 28:250-263(1975).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
DR PIR: A92978; KRGT3J.
DR PIR: B92978; KRGT3M.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin.
FT VARIANT 23 23 Y -> C (IN MINOR COMPONENT).
FT VARIANT 119 119 MISSING (IN MINOR COMPONENT).
FT SEQUENCE 132 AA; 14255 MW; 6C0CC8FA13AB28B CRC64;
SQ
Query Match
Best Local Similarity 19.1%; Score 284; DB 1; Length 132;
Matches 57; Conservative 11; Mismatches 56; Indels 18; Gaps 5;
QY 4 SCGGSVCSDSGGGLGQESCCR-PSCCQTTTCRTTCCRPSCCISCCRPSCCISCCRP 62
DB 3 SCGGPTFSSLSGCGGCLQPRYYRDPCCCRPVSCQTTTTSRPTVTSRTRP-----ICBP 56
QY 63 SCLATTCCTTCCTCRPSCCISCCRP-SCCISSC-----CKPSCCTTCCRPSCCISCCRP 117
DB 57 -----CRPVCCDPCSLQEGGCRPTTCPTSCQAVVCRPCCWATTCOPVSVQPCCRP 110
QY 118 SCISCCCKPSCRTTCCRPSC 139
DB 111 TSCQAPACSRRTTCRTFTSPCC 132
RESULT 9
KR3A_SHEEP
ID KR3A_SHEEP STANDARD; PRT: 130 AA.
AC P02443;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Keratin, high-sulfur matrix protein, IIA3A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RC MEDLINE=74022242; PubMed=4584026;
RA Swart L.S., Haylett T.;
RT "Studies on the high-sulphur proteins of reduced Merino wool. Amino
RT acid sequence of protein SCMB-3A3.";
RL Biochem. J. 133:641-654(1973).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
CC -!- MISCELLANEOUS: THE SOURCE OF THIS PROTEIN IS MERINO WOOL..
DR PIR: A02840; KRSHA3.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin.
SQ SEQUENCE 131 AA; 14163 MW; 8C0B56C0070B1C50 CRC64;
Query Match
Best Local Similarity 18.2%; Score 269.5; DB 1; Length 131;
Matches 54; Conservative 12; Mismatches 58; Indels 17; Gaps 6;
QY 4 SCGGSVCSDSGGGLGQESCCR-PSCCQTTTCRTTCCRPSCCISCCRPSCCISCCRP 63
DB 3 SCGGPTFSSLSGCGG-----CLQPRYYRDP-----CCRPVSC-OTVSRPTVTSRTRP 52
QY 64 CCLTTCCTTCCTCRPSCCISCCRP-SCCISSC-----CKPSCCTTCCRPSCCISCCRP 118
DB 53 C---EPCRRPVCCDPCSLQEGGCRPTTCPTSCQAVVCRPCCWATTCOPVSVQTCRPT 110
QY 119 CCISCCCKPSCRTTTCRPSC 139
| : : : | | | | |
```

```
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
CC -!- MISCELLANEOUS: THE SOURCE OF THIS KERATIN IS MERINO WOOL..
DR PIR: A02841; KRSH3A.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin.
SQ SEQUENCE 130 AA; 13894 MW; 8ACF3719FBBB8361 CRC64;
Query Match
Best Local Similarity 18.8%; Score 279; DB 1; Length 130;
Matches 58; Conservative 11; Mismatches 54; Indels 24; Gaps 7;
QY 4 SCGGSVCSDSGGGLGQESCCR-PSCCQTTTCRTTCCRPSCCISCCRPSCCISCCRP 63
DB 3 SCGGPTFSSLSGCGG-----CLQPCYRDP-----CCRPVSTQTVSRPTVTSRTRP 53
QY 64 CCLTTCCTTCCTCRPSCCISCCRP-SCCISSC-----CKPSCCTTCCRPSCCISCCRP 118
DB 54 C---EPCRRPVCCDPCSLQEGGCRPTTCPTSCQAVVCRPCCWATTCOPVSVQTCRPT 111
QY 119 CCISCCCKPSCRTTTCRPSCCISCC 145
DB 112 SC-----PSAPKPTTC--KTFRTSPCC 130
RESULT 10
KR3A_SHEEP
ID KR3A_SHEEP STANDARD; PRT: 131 AA.
AC P02441;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-AUG-1991 (Rel. 19, Last annotation update)
DE Keratin, high-sulfur matrix protein, IIA3A.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE.
RC MEDLINE=74022242; PubMed=4584026;
RA Swart L.S., Haylett T.;
RT "Studies on the high-sulphur proteins of reduced Merino wool. Amino
RT acid sequence of protein SCMB-3A3.";
RL Biochem. J. 133:641-654(1973).
CC -!- FUNCTION: THE KERATIN PRODUCTS OF MAMMALIAN EPIDERMAL DERIVATIVES
CC SUCH AS WOOL AND HAIR CONSIST OF MICROFIBRILS EMBEDDED IN A RIGID
CC MATRIX OF OTHER PROTEINS. THE MATRIX PROTEINS INCLUDE THE HIGH-
CC SULFUR AND HIGH-TYROSINE KERATINS, HAVING MOLECULAR WEIGHTS OF
CC 6-20 kDa, WHEREAS THE MICROFIBRILS CONTAIN THE LARGER, LOW-SULFUR
CC KERATINS (40-56 kDa).
CC -!- MISCELLANEOUS: THE SOURCE OF THIS PROTEIN IS MERINO WOOL..
DR PIR: A02840; KRSHA3.
DR InterPro: IPR002494; Keratin_B2.
DR Pfam: PF01500; Keratin_B2; 1.
KW Keratin.
SQ SEQUENCE 131 AA; 14163 MW; 8C0B56C0070B1C50 CRC64;
Query Match
Best Local Similarity 18.2%; Score 269.5; DB 1; Length 131;
Matches 54; Conservative 12; Mismatches 58; Indels 17; Gaps 6;
QY 4 SCGGSVCSDSGGGLGQESCCR-PSCCQTTTCRTTCCRPSCCISCCRPSCCISCCRP 63
DB 3 SCGGPTFSSLSGCGG-----CLQPRYYRDP-----CCRPVSC-OTVSRPTVTSRTRP 52
QY 64 CCLTTCCTTCCTCRPSCCISCCRP-SCCISSC-----CKPSCCTTCCRPSCCISCCRP 118
DB 53 C---EPCRRPVCCDPCSLQEGGCRPTTCPTSCQAVVCRPCCWATTCOPVSVQTCRPT 110
QY 119 CCISCCCKPSCRTTTCRPSC 139
| : : : | | | | |
```

```
Db 111 SCOPAPCSRTTCRTFTSPCC 131

RESULT 11
YAV2_XANCV
ID YAV2_XANCV STANDARD; PRT; 784 AA.
AC P14728;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 82 kDa avirulence protein in AVRBS3 region.
OS Xanthomonas campestris (pv. vesicatoria).
OG Plasmid pXv11.
OC Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate 71-21;
RX MEDLINE=89384426; PubMed=2550761;
RA Bonas U., Stall R.E., Staskiewicz B.;
RT *Genetic and structural characterization of the avirulence gene
RT avrBs3 from Xanthomonas campestris pv. vesicatoria";
RL Mol. Gen. Genet. 218:127-136(1989)
CC -!- DISEASE: THE TERM AVIRULENT DESCRIBES A POTENTIALLY VIRULENT
CC PATHOGEN WHICH IS UNABLE TO INDUCE DISEASE SYMPTOMS IN SPECIFIC
CC CULTIVARS OF PEPPER.
CC -!- MISCELLANEOUS: THIS IS ONE OF THE HYPOTHETICAL PROTEINS CODED BY
CC THE OPEN READING FRAMES WITHIN THE REGION REQUIRED FOR AVRBS3
CC ACTIVITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X16130; ; NOT_ANNOTATED_CDS.
CC DR PIR; JQ0317; JQ0317.
CC DR InterPro; IPR005042; Avirulence.
CC DR Pfam; PF03377; Avirulence; 1.
CC DR Hypothetical protein; Plasmid; Virulence.
CC SQ SEQUENCE 784 AA; 82074 MW; 3B2624B411793744 CRC64;

Query Match 15.1%; Score 224.5; DB 1; Length 784;
Best Local Similarity 22.6%; Pred. No. 9.1e-12;
Matches 109; Conservative 19; Mismatches 88; Indels 267; Gaps 29;

QY 1 MVSSCCG-----SVCDSQCGQLGQESCCRP--CCQTTCC----- 34
DB 121 MATTCGVRPWAHSTGNSRCTVSS-----ACLPPLSLAMATTCGVRPWAHSTGN 172
QY 35 -----CRITCC--RP-----SCCISCCRP----- 52
DB 173 SRTVTSAGLPPLLAMATTCGVRPWAHSTGNSRCTVSSACLPPLSLAMATTCGVRP 232
QY 53 -----SCCISSCCKPSCCLTTCRTTC--CRP-----SCCISCCRP----- 87
DB 233 WAWHSTGNSRCTVSSACLPPLSLAMA--TTTCGVRPWAHSTGNSRCTVSSACLPPLSLA 290
QY 88 -----SCCISSCCKPS--CCRTTC--CRP-----SCCISCC 114
DB 291 MATTCGVRPWAHSTGNSRCTVSSACLPPLLAMATTCGVRPWAHSTGNSACTVSSA 350
QY 115 CRP-----SCCISSCCKPS--CCRTTC--RP----- 137
DB 351 CLPPLLAMATTCGVRPWAHSTGNSRCTVSSACLPPLLAMATTCGVRPWAHSTGN 410
QY 138 -SCCISSCCRP-----SCCISSCCKPS--CCRTTC--CRP 167
DB 411 SRTVTSAGLPPLSLAMATTCGVRPWAHSTGNSRCTVSSACLPPLSLAMATTCGVRP 470
```

```
QY 168 -----SCCISSCYRPOCCOP-----SCCRP-----ACCISSCCHP----- 197
DB 471 WAWHSTGNSACTVSSA-----CLPPLLAMATTCGVRPWAHSTGNSACTVSSACLPPL 526
QY 198 -----SCCVSSCRCP-----FSCPTTC--CRITCFHP1-----CCGS 227
DB 527 LAMATTCGVRPWAHSTGNSACTVSSACLPPLLAMATTCGVRPWAHSTGNSRCTVSS 586
QY 228 SCC 230
DB 587 SAC 589

RESULT 12
MCS_MOUSE STANDARD; PRT; 197 AA.
ID MCS_MOUSE
AC P15265;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sperm mitochondrial capsule selenoprotein (MCS).
GN MCSP OR MCS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93039675; PubMed=1418626;
RA Karimpour I., Cutler M., Shih D., Smith J., Kleene K.C.;
RT "Sequence of the gene encoding the mitochondrial capsule
RT selenocysteine codons.";
RL Dev. Biol. 137:395-402(1990).
RN [2]
RP SEQUENCE OF 55-197 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=90152148; PubMed=2303168;
RA Kleene K.C., Smith J., Ruzsaszadeh A., Harris M., Hahn L.,
RA Karimpour I., Gerstel J.;
RT "Sequence and developmental expression of the mRNA encoding the
RT seleno-protein of the sperm mitochondrial capsule in the mouse.";
RL Dev. Biol. 137:395-402(1990).
CC -!- FUNCTION: STRUCTURAL PROTEIN OF THE SPERM MITOCHONDRIAL CAPSULE.
CC IMPORTANT FOR THE MAINTENANCE AND STABILIZATION OF THE CRESCENT
CC STRUCTURE OF THE SPERM MITOCHONDRIA.
CC -!- SUBCELLULAR LOCATION: KERATINOUS MITOCHONDRIAL CAPSULE.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88463; AAB08438.1; -.
CC DR EMBL; M29603; AAA53045.1; -.
CC DR PIR; A37199; A37199.
CC DR MGI; MGI:96945; Mcsp.
CC Mitochondrion; Selenium; Selenocysteine; Sperm; Testis;
CC Spermatoogenesis.
FT SE_CYS 7
FT SE_CYS 17
FT SE_CYS 34
FT SE_CYS 34
SQ SEQUENCE 197 AA; 21015 MW; 721884F7129F02B1 CRC64;

Query Match 14.9%; Score 221.5; DB 1; Length 197;
Best Local Similarity 31.9%; Pred. No. 7.2e-12;
```


DE Shuttle craft protein.
GN STC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=96104568; PubMed=8524296;
RA Strumbakis N.D., Li Z., Tolia P.P.;
RT "A homolog of human transcription factor NF-X1 encoded by the
RT Drosophila shuttle craft gene is required in the embryonic central
RT nervous system.";
RL Mol. Cell. Biol. 16:192-201(1996).
CC -|- FUNCTION: PLAYS AN ESSENTIAL ROLE DURING THE LATE STAGES OF
CC EMBRYONIC NEUROGENESIS. MAY EITHER FINE-TUNE THE GUIDANCE OF THE
CC SPATIAL MAINTENANCE OF THE MIGRATING SNB AND IN NERVE ROOTS,
CC WHICH ARE COMPOSED OF AXONS ORIGINATING FROM DISTINCT GROUPS OF
CC MOTOR NEURONS AND MAY BE REQUIRED TO EITHER GUIDE OR MAINTAIN THE
CC POSITION OF THESE NERVES ALONG A DIRECT AND STRAIGHT PATH TO THEIR
CC ULTIMATE TARGETS IN PARTICULAR MUSCLE FIELDS. MAY PLAY A ROLE IN
CC EGG CHAMBER DEVELOPMENT AND/OR MAY CONFER ESSENTIAL MATERNAL
CC CONTRIBUTIONS TO THE EARLY EMBRYO.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- TISSUE SPECIFICITY: OVARIES AND EMBRYONIC CENTRAL NERVOUS SYSTEM.
CC -|- DEVELOPMENTAL STAGE: MAJOR EXPRESSION IS SEEN IN THE OVARIES WHILE
CC MODERATE LEVELS OF EXPRESSION ARE OBSERVED DURING EMBRYOGENESIS
CC AND THROUGHOUT SUBSEQUENT STAGES OF FLY DEVELOPMENT.
CC -|- SIMILARITY: TO YEAST YNL023C AND HUMAN NF-X1.
CC -|- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U09306; AAB60255.1; -
CC TRANSFAC: T01688; -
CC FlyBase: FBgn0001978; stc.
CC InterPro: IPR001374; R3H.
CC DR InterPro: IPR000967; Znf_NFX1.
CC DR InterPro: IPR001841; Znf_Ring.
CC DR Pfam: PF01422; zf-NF-X1; 8.
CC DR Pfam: PF01424; R3H; 1.
CC DR SMART: SM00393; R3H; 1.
CC DR SMART: SM00184; RING; 1.
CC DR SMART: SM00438; Znf_NFX; 9.
CC DR PROSITE: PS01359; ZF_PHD.1; 1.
CC DR PROSITE: PS00518; ZF_RING.1; 1.
CC DR PROSITE: PS50089; ZF_RING.2; 1.
CC DR Transcription regulation: DNA-binding; Nuclear protein; Repeat;
CC KW Alternative splicing; RNA-binding; Zinc-finger.
CC FT DOMAIN 241 268
CC FT TANDEN REPEATS OF R-D.
CC FT ZN_FING 386 433
CC FT RING-TYPE.
CC FT DOMAIN 464 904
CC FT 7 X APPROXIMATE REPEATS, CYS-RICH.
CC FT REPEAT 464 500
CC FT REPEAT 517 554
CC FT REPEAT 575 614
CC FT REPEAT 634 675
CC FT REPEAT 723 762
CC FT REPEAT 834 871
CC FT REPEAT 875 904
CC FT VARSPLIC 109 115
CC FT MISSING (IN SHORT ISOFORM).
CC SQ SEQUENCE 1106 AA; 123261 MW; 8CE146DB37BB8D65 CRC64;

Query Match 14.2%; Score 210; DB 1; Length 1106;
Best Local Similarity 22.4%; Pred. No. 1.6e-10;
Matches 76; Conservative 22; Mismatches 120; Indels 122; Gaps 15;

QY 6 CGSVCSQD-QSCGGGLGQESCCRPSC-----QTTCKR 36
DB 517 CGEIKLLKNGEHRQ-QAFCHSGKCAACSEVVQQCHGCKQERKVPCTRESQURKTYSK 576
QY 37 TTCCRPSCCJSSCCRPSCCJSSC--CKPSCCLTTCC-----RTTCCRPV-SGCLIS 82
DB 577 DSCGQPLPCGHHKCKDSCHAGSCRPCKLSPQITSCPCGKIPVPAGORSCLDIPITCKG 636
QY 83 SSCR-----PSCCJSSC-----CKPSC-----C 100
DB 637 ICSRTLRCGKPAHPHQVGSCKHLGOCPPCKOTGVKRCGCHMDOMIKKCOLCNRAADARC 696
QY 101 RTTCCRPSCCJSSCCHPSCC--ISSCKPSCCCTTCRPSCCJSSCCRPSC--CISCK 156
DB 697 KRCTKRRSGKHKVNECCIDIDHHCPLPCNRTLSGKHKCDQPCIRGNCPICYRSSFE 756
QY 157 PSCCOTTCCRPSCCJSSCYHQQC--ONSCCRPACCJSSC-----CH-----PSCKV 201
DB 757 ELYCE-----CGAEVYHPVPCGKPKICKLPSSRIHPCDHPPOHCHSGITCTTCMI 809
QY 202 SSCR-----CPFSCTTCTTCRTTCRPHICCGSSCC 230
DB 810 FTTKLCGHNHWRKTIPTCSUPNISCMAUCPKPLPCGGHKC 849

RESULT 15
NFX1_HUMAN
ID NFX1_HUMAN STANDARD; PRT; 1104 AA.
AC Q12986;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional repressor NF-X1 (Nuclear Transcription factor, X box-
DE binding, 1).
GN NFX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95053707; PubMed=7964459;
RA Song Z., Krishna S., Thanos D., Strominger J.L., Ono S.J.;
RT "A novel cysteine-rich sequence-specific DNA-binding protein
RT interacts with the conserved X-box motif of the human major
RT histocompatibility complex class II genes via a repeated Cys-His
RT domain and functions as a transcriptional repressor.";
RL J. Exp. Med. 180:1763-1774(1994).
CC -|- FUNCTION: REPRESSOR OF HLA-DRA TRANSCRIPTION. BINDS TO THE X-BOX
CC MOTIF OF CLASS II MHC GENES. MAY PLAY AN IMPORTANT ROLE IN
CC REGULATING THE DURATION OF AN INFLAMMATORY RESPONSE BY LIMITING
CC THE PERIOD IN WHICH CLASS II MHC MOLECULES ARE INDUCED BY
CC INTERFERON-GAMMA.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- INDUCTION: BY INTERFERON GAMMA.
CC -|- SIMILARITY: TO D.MELANOGASTER SHUTTLE CRAFT PROTEIN (STC) AND
CC YEAST YNL023C.
CC -|- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U15306; AAA69517.1;
CC TRANSFAC: T01694; -
CC Genew: HGNC:7803; NFX1.
CC DR MIM: 603255; -
CC DR InterPro: IPR001374; R3H.

```
DR InterPro: IPR000967; Znf_NFX1.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_Ring.
DR Pfam: PF01422; zf-NF-X1; 8.
DR SMART: SM01424; R3H; 1.
DR SMART: SM00393; R3H; 1.
DR SMART: SM00184; RING; 1.
DR SMART: SM00438; Znf_NFX; 9.
DR PROSITE: PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE: PS00016; ZF_PHD_2; 1.
DR PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
DR PROSITE: PS00089; ZF_RING_2; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Repeat; Zinc-finger.
FT ZN_FING 342 393 RING-TYPE.
FT DOMAIN 424 876 7 X APPROXIMATE REPEATS, CYS-RICH.
FT REPEAT 424 463 1.
FT REPEAT 480 517 2.
FT REPEAT 541 580 3.
FT REPEAT 606 647 4.
FT REPEAT 695 734 5.
FT REPEAT 806 842 6.
FT REPEAT 843 876 7.
SQ SEQUENCE 1104 AA; 123149 MW; 36FEC4292F78130C CRC64;

Query Match 13.3%; Score 198; DB 1; Length 1104;
Best Local Similarity 21.9%; Pred. No. 1.5e-09;
Matches 80; Conservative 18; Mismatches 118; Indels 150; Gaps 19;

Qy 6 CGSVCSQDSGC-----QGLQDSC-----CRPSCCQ-----TT 33
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 517 CGSTRDVLCTGVKSDGDFSLCTGKDLKCGNHTCSQVCHPOPCOOCPLQLVR 576
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 34 CC-----RTTCRP-SCCISCCRPSC-----ISSC-----CKP 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 577 CCPCGQTPLSOLLELGLSSSRKTCMDVPVPSGKVGKPLPCGSLDFIHTCEKLCHEGDCGP 636
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 63 -----SC-----C-----LTTCRTTCCRPSCCISCCRPSCCIS--CCK 96
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 VSRTSVISRCGSFRTKELPCTSLKSEDATFMDKRCNKNKRLCGRIKNEICCVDEHKCP 696
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 97 PSCGRTTCCRPSCCISCCRPSCCISCCCKPSCCCTTC-----CRPSCCIS 142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 LNCGRKURCGLHRCCEPCHRGNC--QTCWQASFDLTCGASVIYPPVPCGTRPPECTQ 754
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 143 SCCR-----PSC-----CIS-----SCCRPSCCOT 162
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 755 TCARVHEDHPVYHSGHSEKCPCTFLTQKWMGKHEFRSNIPCHLVDISGLP-CSAT 813
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 163 TCCRPSCCISCYRPQCCQPCSCRPACCI--SSCHPSCCVSSCRCPFCPTTCRTTCF 220
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 814 LPCGMHKCQRLCHGCELVDEPCQPCCTTPRADCGHP--CNAPCHTSSPCPVACKAKVE 871
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 221 HPICCG 226
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 872 LQCECG 877
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: June 11, 2003, 11:04:58

Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 11:02:42 ; Search time 72 Seconds
(without alignments)
425.662 Million cell

Title: US-09-874-062-3
 Perfect score: 1484
 Sequence: 1 MYSSCCGSVCSDQSCGQLG.....PTTCRTTCFHPICCGSSCC 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID52/cgcdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID52/cgcdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID52/cgcdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID52/cgcdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1484	100.0	230	22	AAE02058
2	1218.5	82.1	206	22	ABG29388
3	767	51.7	301	22	AAW79986
4	758.5	51.1	250	22	AAW79002
5	599	40.4	120	22	ABBI12277
6	552.5	37.2	256	22	ABG08337
7	552.5	37.2	351	22	ABG23698
8	523.5	35.3	267	22	ABG27981
9	513	34.6	231	22	ABB32110
10	513	34.6	231	22	ABB37362
					Human four disulfide
					Novel human dihydro
					Human protein SEQ
					Human protein SEQ
					Human hair keratin
					Novel human dihydro
					Novel human dihydro
					Peptide #4761 enco
					Peptide #4868 enco

ALIGNMENTS

RESULT 1
AAE02058
ID AAE02058 standard; Protein; 230 AA.

AC AAE02058;

DT 10-AUG-2001 (first entry)

Human four disulfide core domain (FDCD)-containing protein.

KW	Human; four disulfide core domain; FDCD; immunosuppressive; cytostatic;
KW	antiarthritic; antirheumatic; antiproliferative; cardiant; vasotropic;
KW	carcinoprotective; nootropic; neuroprotective; antibacterial; virucide;
KW	fungicide; ophthalmological; dermatological; antiinflammatory; vulnerary;
KW	antiarteriosclerotic; gene therapy; skin disorder; congenital disorder;
KW	mole; freckle; haemangioma; integumentary tumour; basal cell carcinoma;
KW	keratosis; melanoma; atherosclerosis; urticaria; photosensitivity;
KW	eczema; autoimmune disorder; lupus erythematosus; scleroderma; keloid;
KW	stria; erythema; petechia; purpura; xanthelasma; autoimmune disease;
KW	rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW	cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW	neoplasm; angiodysplasia; nervous system disorder; Alzheimer's disease;
KW	infection; ocular disorder; corneal infection.

OS Homo sapiens.

PN WO200140249-A1.

07-JUN-2001.

29-NOV-2000; 2000WO-US32462

PR 01-DEC-1999: 99US-0168229

protein #4647 enco
Human brain express
Human bone marrow
Peptide #4762 enco
Peptide #4852 enco
Peptide #4852 enco
Peptide #4617 enco
Human peptide enco
Human peptide enco
Peptide #7027 enco
Protein #6285 enco
Human brain express
Human bone marrow
Peptide #6193 enco
Peptide #7105 enco
Human peptide enco
Human peptide enco
Peptide #6312 enco
Human brain express
Human bone marrow
Peptide #6315 enco
Human peptide enco
Human protein Sg
Sheep keratin-asso
Novel human diagno
Human diagno
Human NgW3a protei
Novel human diagno
Human polypeptide
Human polypeptide
1-aminocyclopropan
Peptide #7583 enco
Peptide #7583 enco
Protein #6565 enco
Human brain express
Human bone marrow
Peptide #6429 enco
Peptide #7747 enco
Human peptide enco
Human peptide enco
Human epidermal pr

XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Shi Y;
PI WPI: 2001-356153/37.
XX N-PSDB: AAD06303.
DR
XX
XX
PT Four disulfide core domains (FDCD) containing polypeptide and its
PT polynucleotide are used to prevent, treat or ameliorate a medical
PT condition associated with FDCD e.g. skin disorders -
XX
XX
PS Claim 11: Page 230; 235pp; English.
XX
XX The present sequence is four disulfide core domain (FDCD)-
CC containing protein from human. FDCD DNA and protein are used
CC to prevent, treat or ameliorate skin disorders such as congenital
CC disorders e.g. moles, freckles and haemangiomas as well as
CC integumentary tumours (keratosis, basal cell carcinoma, malignant
CC melanoma), injuries and inflammation of the skin, atherosclerosis,
CC urticaria, eczema, photosensitivity, autoimmune disorders (lupus
CC erythematosus, scleroderma), keloids, striae, erythema, petechiae,
CC purpura and xanthelasma. They are also used in the diagnosis and
CC treatment of diseases and disorders of hair and epithelium,
CC autoimmune diseases (e.g. rheumatoid arthritis), hyperproliferative
CC disorders (e.g. neoplasms of the breast or liver), cardiovascular
CC disorders (e.g. cardiac arrest), cerebrovascular disorders (e.g.
CC cerebral ischaemia, angiodenesis), nervous system disorders (e.g.
CC Alzheimer's disease), infections caused by bacteria, viruses and
CC fungi and ocular disorders (e.g. corneal infection). The FDCD-
CC containing proteins are used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. They are used in
CC diagnosing a pathological condition or susceptibility to a pathological
CC condition. The antibodies to FDCD protein are used in alleviating
CC symptoms associated with the disorders and in diagnostic immunoassays
CC e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA).
CC The FDCD sequences are also used in gene therapy.
XX
XX Sequence 230 AA:
XX
XX Query Match 100.0%; Score 1484; DB 22; Length 230;
XX Best Local Similarity 100.0%; Pred. No. 2.9e-94;
XX Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVSSCGSVCSQSGGGLGQESCCRPSCCCTTCRTTCRPSCCISCCRPSCCISCC 60
DB 1 MVSSCGSVCSQSGGGLGQESCCRPSCCCTTCRTTCRPSCCISCCRPSCCISCC 60
QY 61 KPSCLTTCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 120
DB 61 KPSCLTTCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 120
QY 121 ISSCKKPSCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 180
DB 121 ISSCKKPSCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 180
QY 181 QPSCCRPACCISSCCHPSCCVSSCRCPSCPTTCRTTCFHPICGSSCC 230
DB 181 QPSCCRPACCISSCCHPSCCVSSCRCPSCPTTCRTTCFHPICGSSCC 230
RESULT 2
ABG29368
ID ABG29368 standard; Protein: 206 AA.
XX
XX AC ABG29368;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #29359.
XX

Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.
Homo sapiens.
WO200175067-A2.
11-OCT-2001.
30-MAR-2001; 2001WO-US08631.
31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HYSE-) HYSEQ INC.
Drmanac RT, Liu C, Tang YT;
WPI: 2001-639362/73.
N-PSDB: AAS93555.
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
Claim 20; SEQ ID NO 59727; 103pp; English.
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG0010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 206 AA:
Query Match 82.1%; Score 1218.5; DB 22; Length 206;
Best Local Similarity 84.3%; Pred. No. 3.2e-76;
Matches 194; Conservative 0; Mismatches 1; Indels 35; Gaps 1;
QY 1 MVSSCGSVCSQSGGGLGQESCCRPSCCCTTCRTTCRPSCCISCCRPSCCISCC 60
DB 12 MVSSCGSVCSQSGGGLGQESCCRPSCCCTTCRTTCRPSCCISCCRPSCCISCC 42
QY 61 KPSCLTTCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 120
DB 43 -----TTCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 96
QY 121 ISSCKKPSCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 180
DB 97 ISSCKKPSCCRTTCRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRPSCCISCCRPSCC 156
QY 181 QPSCCRPACCISSCCHPSCCVSSCRCPSCPTTCRTTCFHPICGSSCC 230
DB 157 QPSCCRPACCISSCCHPSCCVSSCRCPSCPTTCRTTCFHPICGSSCC 206

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB; AAS72524.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 20; SEQ ID No 38696; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 256 AA;

Query Match 37.2%; Score 552.5; DB 22; Length 256;
Best Local Similarity 40.6%; Pred. No. 7.7e-31;
Matches 101; Conservative 29; Mismatches 60; Indels 59; Gaps 11;
QY 9 VCSQSCGGLGQESC-----CRPSCQTTTCRTTCRPPSCCISCCRPSCCIS-- 58
DB 39 VCLPGSC-----DSCSDSNQVDDCPESCCEPPCCAPSCCAPCLTLVCTPVSCVSSPC 92
QY 59 ---CCRPSCCLTTCRTTCRPPSCCISCCRPSCCISCCRPSCCISCCRPSCCISCC 115
DB 93 CQACEPSPCQSGC--TSSCTPSCCQSSCQACCTSSPCQACCPVCCPKVCCPVCC 150
QY 116 RPSCCISSCCKP-----SCRTTCRPPSCCISCCRPSCCISCCRPSCCISCCRPSCC 168
DB 151 KPCCCKPICCPVCSGASSCCQSSQRPACCTTSCRPSSSVSLLCRP-VCRSTCCVP- 208
QY 169 CCISSCVRP--QCQPCRCRPAACISCCCHPSCC-VSSCRCPFCPTTCRTTCFHPICCG 226
DB 209 --IPSCCAPASTCQPCRCRPAACISCCCHPSCC-VSSCRCPFCPTTCRTTCFHPICCG 247
QY 227 -----SSC 229
|||

DB 248 LSSGQKSSC 256

RESULT 7

ABG23698

ID ABG23698 standard; Protein; 351 AA.

XX AC ABG23698;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #23689.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB; AAS87885.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 20; SEQ ID No 54057; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 351 AA;

Query Match 37.2%; Score 552.5; DB 22; Length 351;
Best Local Similarity 42.4%; Pred. No. 9.6e-31;
Matches 97; Conservative 28; Mismatches 59; Indels 45; Gaps 9;
QY 21 QESCCRPSCQTTTCRTTCRPPSCCISCCRPSCCISCCRPSCCIS--CCRPSCCLTTCRTTC 75
DB 148 QVDACPESCGEPCCGATSCCAPACLTLCVTPVSCVSSPCQACEPSCQSGC--TSSC 205

CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 231 AA;

Query Match 34.6%; Score 513; DB 22; Length 231;
 Best Local Similarity 40.6%; Pred. No. 3.5e-28;
 Matches 86; Conservative 29; Mismatches 77; Indels 20; Gaps 3;
 QY 29 CCCTTCCTTCCTCCRPSCCISCCRPSCCISCCRPSCCLTTCCRTTCCTCCRPSCCISCCRPSC 88
 Db 1 CCQVGVSEATSCQPVLCVPTSCQPVLCCKSCCQ-----VCEPSCCSAVCTLPS 50
 QY 89 CCISCCCKPSCCTTCCTCCRPSCCISCCRPSCCISCCCKPSCCTTCCTCCRPSCCIS-----S 143
 Db 51 SCQPVVCEPSCCQPVCTPTCSVTSSCOAVCCDPSPCEPSCSSESSICQATCVALVCEPV 110
 QY 144 CCRPSCCISCCCKPSCCCTTCCTCCRPSCCISCCRPSCCISCCRPSCCISCCCKPSC----- 199
 Db 111 CLRPVCCVQSSCEPPSPVSTCQEPSCCVSSICQEPSCSPSPSPSPSPSPSPSPSPSPSPSPSP 170
 QY 200 -CVSSCRCPSCPTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 230
 Db 171 RCPSPVCEPVPSCPTSCRPLSCSPGSSASATC 202

RESULT 10
 ARB37362
 ID ARB37362 standard; Peptide; 231 AA.
 XX
 AC AHB37362;
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #4868 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human foetal liver -
 XX
 PS Claim 27; SEQ ID NO 29997; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 231 AA;

Query Match 34.6%; Score 513; DB 22; Length 231;
 Best Local Similarity 40.6%; Pred. No. 3.5e-28;
 Matches 86; Conservative 29; Mismatches 77; Indels 20; Gaps 3;
 QY 29 CCCTTCCTTCCTCCRPSCCISCCRPSCCISCCRPSCCLTTCCRTTCCTCCRPSCCISCCRPSC 88
 Db 1 CCQVGVSEATSCQPVLCVPTSCQPVLCCKSCCQ-----VCEPSCCSAVCTLPS 50
 QY 89 CCISCCCKPSCCTTCCTCCRPSCCISCCRPSCCISCCCKPSCCTTCCTCCRPSCCIS-----S 143
 Db 51 SCQPVVCEPSCCQPVCTPTCSVTSSCOAVCCDPSPCEPSCSSESSICQATCVALVCEPV 110
 QY 144 CCRPSCCISCCCKPSCCCTTCCTCCRPSCCISCCRPSCCISCCRPSCCISCCCKPSC----- 199
 Db 111 CLRPVCCVQSSCEPPSPVSTCQEPSCCVSSICQEPSCSPSPSPSPSPSPSPSPSPSPSPSPSP 170
 QY 200 -CVSSCRCPSCPTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 230
 Db 171 RCPSPVCEPVPSCPTSCRPLSCSPGSSASATC 202

RESULT 11
 ARB22648
 ID ARB22648 standard; Protein; 231 AA.
 XX
 AC ABB22648;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #4647 encoded by probe for measuring heart cell gene expression.
 XX
 KW Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200157274-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488899/53.
 XX

DR WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow.
XX Example 4; SEQ ID NO: 30798; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX Sequence 231 AA;
SQ
Query Match 34.6%; Score 513; DB 22; Length 231;
Best Local Similarity 40.6%; Pred. NO. 3.5e-28;
Matches 86; Conservative 29; Mismatches 77; Indels 20; Gaps 3;
QY 29 CCQTTCCRTTCRPSGCCISCCRPSCGCISSCKPSCCLTTCRTTCCRPSCGCISSCCRPSS 88
DB 1 CCQPVGSEATSCQPVLCVPTSCQPVLCSSCCQP-----VWCEPSCSAVCTLPS 50
QY 89 CCISSCCKPSCCRTTCCRPSCGCISSCCRPSCGCISSCCRPSCGCISSCCRPSCGCISSCCRPSS 143
DB 51 SCQPVCEPSCCQPVCTPTTCGVTSSCAVCCDPSCEPSCSESSICQATCVALVCEPV 110
QY 144 CCRPSGCCISCCKPSCCGTCCTCRPSGCCISCCYRPOCCPSCCRPACGCISSCCRPSC----- 199
DB 111 CLRPVCCVQSCEPSPVSTQEPSCGCVSSICQPTICSEPSAPVCVSSPCQPTCYVVK 170
QY 200 -CVSSCRCPSCPTTCRTTCFHPICCGSSCC 230
DB 171 RCPVCPPEPVPSTSCRLSCSPGSSASAIC 202
RESULT 14
AAM18328
ID AAM18328 standard; Protein; 231 AA.
XX
AC AAM18328;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #4762 encoded by probe for measuring cervical gene expression.
XX
KW Probe: human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX Homo sapiens.
XX OS
XX WO200157278-A2.
XX
XX PD
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -
XX Claim 27; SEQ ID NO 23154; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 231 AA;
SQ
Query Match 34.6%; Score 513; DB 22; Length 231;
Best Local Similarity 40.6%; Pred. NO. 3.5e-28;
Matches 86; Conservative 29; Mismatches 77; Indels 20; Gaps 3;
QY 29 CCQTTCCRTTCRPSGCCISCCRPSCGCISSCKPSCCLTTCRTTCCRPSCGCISSCCRPSS 88
DB 1 CCQPVGSEATSCQPVLCVPTSCQPVLCSSCCQP-----VWCEPSCSAVCTLPS 50
QY 89 CCISSCCKPSCCRTTCCRPSCGCISSCCRPSCGCISSCCRPSCGCISSCCRPSCGCISSCCRPSS 143
DB 51 SCQPVCEPSCCQPVCTPTTCGVTSSCAVCCDPSCEPSCSESSICQATCVALVCEPV 110
QY 144 CCRPSGCCISCCKPSCCGTCCTCRPSGCCISCCYRPOCCPSCCRPACGCISSCCRPSC----- 199
DB 111 CLRPVCCVQSCEPSPVSTQEPSCGCVSSICQPTICSEPSAPVCVSSPCQPTCYVVK 170
QY 200 -CVSSCRCPSCPTTCRTTCFHPICCGSSCC 230
DB 171 RCPVCPPEPVPSTSCRLSCSPGSSASAIC 202
RESULT 15
AAM30815
ID AAM30815 standard; Protein; 231 AA.
XX
AC AAM30815;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #4852 encoded by probe for measuring placental gene expression.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
XX OS Homo sapiens.
XX
XX WO200157272-A2.
XX
XX PD
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for

[illegible]

Search completed: June 11, 2003, 11:04:04
Job time : 74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Kun on: June 11, 2003, 11:02:46 ; Search time 27 Seconds
(without alignments)
250.640 Million cell

Title: US-09-874-062-3

perfect score: 1484
Sequence: 1 MVSSCCGVCSDQSCQGLG.....PTTCRTTCFHPICCGSSCC 230

Scoring table: BLOSUM62

Learning curve: 2000 Hz
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length:	2000000000
Maximum DB seq length:	2000000000

Post-processing: Minimum Match 0%

Database : Issued Patents AA:*

```
1: /cgn2_6/pdata/1/iaa/5A_COMB pep: *
2: /cgn2_6/pdata/1/iaa/5B_COMB pep: *
3: /cgn2_6/pdata/1/iaa/6A_COMB pep: *
4: /cgn2_6/pdata/1/iaa/6B_COMB pep: *
5: /cgn2_6/pdata/1/iaa/pctrys_COMB pep: *
6: /cgn2_6/pdata/1/iaa/pctrys_COMB pep: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	412	27.8	1417	4	US-08-900-230-3	Sequence 3, Appli
2	412	27.8	2211	4	US-09-738-884-1	Sequence 1, Appli
3	391.5	26.4	1345	2	US-08-977-767-3	Sequence 3, Appli
4	378	25.5	1400	4	US-08-630-915A-37	Sequence 37, Appli
5	365	24.6	1917	4	US-09-627-650B-5	Sequence 5, Appli
6	365	24.6	1917	4	US-09-436-063C-5	Sequence 5, Appli
7	338.5	22.8	2088	4	US-09-548-372D-13	Sequence 13, Appli
8	338.5	22.8	2088	4	US-09-548-367D-13	Sequence 13, Appli
9	316.5	21.3	1128	4	US-09-627-650B-11	Sequence 11, Appli
10	316.5	21.3	1128	4	US-09-436-063C-11	Sequence 11, Appli
11	316.5	21.3	1652	4	US-09-627-650B-1	Sequence 1, Appli
12	316.5	21.3	1652	4	US-09-436-063C-1	Sequence 1, Appli
13	316.5	21.3	2508	4	US-09-627-650B-7	Sequence 7, Appli
14	316.5	21.3	2508	4	US-09-436-063C-7	Sequence 7, Appli
15	316.5	21.3	2544	4	US-09-627-650B-3	Sequence 3, Appli
16	316.5	21.3	2544	4	US-09-436-063C-3	Sequence 3, Appli
17	316.5	21.3	2601	4	US-09-627-650B-9	Sequence 9, Appli
18	316.5	21.3	2601	4	US-09-436-063C-9	Sequence 9, Appli
19	282	19.0	801	1	US-07-906-349A-6	Sequence 6, Appli
20	209.5	14.1	430	4	US-09-156-836B-2	Sequence 2, Appli
21	204	13.7	430	3	US-08-997-897-2	Sequence 2, Appli
22	198	13.3	1104	2	US-08-327-832-5	Sequence 5, Appli
23	198	13.3	1104	2	US-08-828-584-5	Sequence 5, Appli
24	184.5	12.4	120	3	US-08-508-761B-22	Sequence 22, Appli
25	180.5	12.2	2556	1	US-08-083-590A-20	Sequence 20, Appli
26	180.5	12.2	2556	3	US-08-532-384-20	Sequence 20, Appli
27	179.5	12.1	1443	4	US-08-990-823-112	Sequence 112, App

28	176.5	11.9	2556	1	US-08-185-432-17	Sequence 17, Appl
29	176.5	11.9	2556	4	US-08-899-232-2	Sequence 2, Appl
30	169	11.4	578	4	US-08-981-332-13	Sequence 13, Appl
31	169	11.4	1964	4	US-09-467-997-1	Sequence 1, Appl
32	169	11.4	2050	2	US-08-347-594-2	Sequence 2, Appl
33	162	10.9	109	2	US-08-527-044-2	Sequence 2, Appl
34	162	10.9	109	3	US-09-013-780-2	Sequence 2, Appl
35	162	10.9	341	2	US-08-209-521-11	Sequence 11, Appl
36	162	10.9	589	1	US-07-668-648-2	Sequence 2, Appl
37	162	10.9	589	2	US-08-429-998-2	Sequence 2, Appl
38	162	10.9	589	2	US-08-431-333-2	Sequence 2, Appl
39	162	10.9	589	5	PCT-US91-02321-2	Sequence 2, Appl
40	162	10.9	1810	5	PCT-US95-11684-2	Sequence 4, Appl
41	160.5	10.8	2199	5	PCT-US95-11684-2	Sequence 2, Appl
42	159	10.7	2471	1	US-08-185-432-16	Sequence 16, Appl
43	159	10.7	2471	1	US-08-083-590A-19	Sequence 19, Appl
44	159	10.7	2471	3	US-08-532-384-19	Sequence 19, Appl
45	159	10.7	2471	4	US-08-899-232-1	Sequence 1, Appl

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 27.8%; Score 412; DB 4; Length 1417;
Best Local Similarity 33.1%; Pred. NO. 8.8e-24;
Matches 92; Conservative 13; Mismatches 97; Indels 76; Gaps 22;

5 CCGSVCSNOSCGGL-----GOESCCRPSCC-----TTCCRTT----- 38

Db 899 CCGCGCCTCCTACGCGCTCTGCTGGGTCCGCACCGGCTCATCTGTGCTTCTGGTACG 958

Qy 39 -CCRPSCCISSCOR--PSCCISSCCKPSCCLTTCRTTCRRPSCCISSCCRPSCCISSCC 95
Db 959 GCC--GCTTCCGCTTCAGCCGGCC--ACCTACGCCCTGCC--GCTGGCTTCACACATGCC 1012
Qy 96 KPSCCRTTCRRPSCCISSCCRPSCCISSCCKPSCCCTTC--CRPSCCISSC-C 145
Db 1013 TGGCC--TAC--GCCAACTCTGCTCAAC--CCGCTCGTCTACGGCTCGCTCGGC 1065
Qy 146 RPSCCISSCCKPSC--CQTTC--RPSCCISSCYRPOCCOPSPCRP-ACC-IS 192
Db 1066 ---CACTTCCGGGGCG--TTCCCGCGCTGTGGCGTGC--GGCC--GCCAGCGCGC 1117
Qy 193 SCCHPSCVSCRRPSCPTTCRTTCFHPITCCGSSCC 230
Db 1118 ACCGTGCCCGCCG--GCCITGC--GTCCGTCCCGCCCC 1152

RESULT 2

US-09-738-884-1
; Sequence 1, Application US/09738884
; Patent No. 6391606
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000849
; CURRENT APPLICATION NUMBER: US/09/738,884
; CURRENT FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2211
; TYPE: PRT
; ORGANISM: Human
US-09-738-884-1

Query Match 27.8%; Score 412; DB 4; Length 2211;

Best Local Similarity 29.9%; Pred. No. 1.2e-23;
Matches 88; Conservative 19; Mismatches 99; Indels 88; Gaps 19;

Qy 3 SSCCGSV-----CSDOSC-----GGLGQESCCR-----PSCQTTTC--CRTTCR 41
Db 973 TCCGCTCGTGGAGCTGACTGCTGGAGGGCCAGGAGGGAGCTCATATCAT 1032
Qy 42 PSCCISSCRRPSCCISSCCKPSC--LTTTCRTTCCHPS-----CCISSCRRPSCC 91
Db 1033 GGCCATAC-----CCTCAC--TCCAAGATTCTCTCCGGGACGTGGCCAAAGCGTGGC 1086
Qy 92 SSCCKPSCRTTCRRPSCCISSCCRPSCCISSCCKPSCRTTCRRPSCCISSCCRPSCC 151
Db 1087 GACCATGCTTCACG---CTGTCC---CCTTACCTGTGC-ATCTATCCCTGGGAGAACCA 1139
Qy 152 SSCCKPSC-----COTTCCRPSCCISSCYRPOCCOPSPCRPACCISSC----- 194
Db 1140 CTGGGGGTGGAGCAGAGGTGCATGGCCGGC---ACC--TCTGCACCATCTGGGG 1194
Qy 195 --CHPSC-----CVSSRCRPSCPTTC--RTTCFHPITCCG-----SSCC 230
Db 1195 GACATGCTGGTGACACAGCGCTGCACTCCCAATC-----CCGAGGAGGTGCC 1244

RESULT 3

US-08-977-767-3
; Sequence 3, Application US/08977767
; Patent No. 5972684
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Greenwald, Sara
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII

; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,767
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0423 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1345 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1532042
US-08-977-767-3

Query Match 26.4%; Score 391.5; DB 2; Length 1345;

Best Local Similarity 30.3%; Pred. No. 2.8e-22;
Matches 86; Conservative 22; Mismatches 105; Indels 71; Gaps 19;

Qy 5 CCG-----SVCSDSGCGGLGQESCCRPSCTTCRTTCRRPSCC1-SSCCRP-----SCC 55
Db 858 CCGCATCTCTTAAAGAAATGATGCC--TACTTTC--TTCAAGACCTGAGCTGAGTGC 913
Qy 56 ISS---GC-KPSCGLTT-----CCTTCRRPSC-CISSCCRPSCCISSCCKPSC- 99
Db 914 TGTGCCCCGAGTCTTTGGCTTCATCACCCTATCAGGGCTCTCTCAGCACCTCAGTTC 973
Qy 100 -----CRFT--CCRPSCCISSCCRPSCCISSCCKPSC-----C 130
Db 974 CGGAGACTGTATACCTGGATCTCATGACAGGGCCCTCAATATCACCTCCCTCAGTGC 1033
Qy 131 RTTCRRPSC-CISSCCRPSC--ISSCKPSC--QTTCCRPSCCISSCYRPOCCPSCCK 186
Db 1034 ATTCCTCCGAGCTTCTGAGCCAGATCC--TCCGGTCCCAGTCTTCAGAGCTCAGTGC 1091
Qy 187 PACCISSCCRPSCVSSRCRPFSCPTTCRTTCFHPITCCGSSCC 230
Db 1092 TAAC-GGCCCGCCCTGCA---GC---CCTCCCGCACAGGGCC 1128

RESULT 4

US-08-630-915A-37
; Sequence 37, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6409820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.


```

; FILE REFERENCE: 21101.000903
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1128
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-11

Query Match          21.3%; Score 316.5; DB 4; Length 1128;
Best Local Similarity 25.1%; Pred. No. 9,2e-17;
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;

QY      4   SCCGVCSDSCGGGLG---QESCRCRPSCCQTTCGRITC-CRPSCCISSCCRSPSCCISCC 59
Db              ||| | : || | | | | | | | | | | | | | | | | | | | | | | |
721 TCGC-CCTTGCCCAANTGCACATCCAGCGGGTCCAACCTCGATCATCGCCGTCAATCAAG 779

QY      60  CKPSCCLLTTCCRTTCCRPSCCIISSC-CRPSCCIISS-----CKKPSCC----RTT 103
Db              | | | | | | | | | | | | | | | | | | | | | | | | | | | |
780 CACTCGAATCGATTCTCTGTGTCAGTCAAGTCACATTACATTCGATCGACGCCGTGCGGGTTTT 839

QY     104  CCR-----PSCCIISSC-CRPSCCIISSCKPSCCRTTCCRPSCCIISSCCRP 147
Db              ||| | : || | | | | | | | | | | | | | | | | | | | | | | |
840 CCTCTGTGTTTCATCTTGTTCAACACTCTCTTCGGCTGATTTC-TACTGTACAAATCCAA 898

QY     148  SC-CISSCC-----KPSCCQNTCCRPSCCIISSCYRPOCCOPSCC 185
Db              | | | | | | | | | | | | | | | | | | | | | | | | | | | |
899 GCCTCTGCCGTATATTAGTGAACACGAGGGTGACCGTTGGCATGC-----TCCAGACC 951

QY     186  RPAC-----CISSCHPSCCVSSCRCPPSC--PTTCCRT-----TCFHPICCG 226
Db              | | : || | | | | | | | | | | | | | | | | | | | | | | |
952 TTCATTAAATCTCAATCC--AACTTCCTCA-TCATTTTCCATTTTCGAATATCTCTTTTCT 1008

QY     227  SSC 229 .
Db              : |
1009 TGC 1011

```

```

RESULT 10
US-09-436-063C-11
; Sequence 11, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095Corrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1128
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-11

```

Query Match	21.3%	Score 316.5	DB 4	Length 1128
Best Local Similarity	25.1%	Pred. No. 9.2e-17		
Matches	76	Conservative	27	Mismatches 111
				Indels 89
				Gaps 17

QY	4	SCGSGVCS	SDSCSGGGLG	--QESCCR	PS	CQTTCRTTC	-CRPSC	CISSCCRPSC	ISSC	59	
										::	
Db	721	TCGG	-CCTTCGGCAAT	CGACAT	CGA	CGCGGCT	CGAACCT	CGATCG	CGCGGCT	CAATCAAG	779

Qy	60	CKPSCCLTTCCTTTCCRPSCC1SSC-CRPSCC1SS-----CCKPSCC--RPT 103
		: : : : : : : : : : : :
Db	780	CAGTCGAATCGATTCTGTCTCAGTCACAGTCACATTGACATCGTCAGCCCGTGCGCGPTT 839
		: : : : : : : : : : : : : :
Qy	104	CCR-----PSCC1SSC-CRPSCC1SSCCCKPSCCTTCRPSCC1SSCGRP 147
		: : : : : : : : : : : : : :
Db	840	CTCTCTCTTTTCATCTCTTCAACACTCTCTCTGGCTGATTC-TACTGTGTAATAATCAA 898
		: : : : : : : : : : : : : :
Qy	148	SC-C1SSCC-----KPSCCCTTCRPSCC1SSCKPSCCTTCRPSCC1SSCGRPSCC 185
		: : : : : : : : : : : : : :
Db	899	GCCTCTGCCCTATATTATGTCACACGAGGCGTGACCGTTGCGATCC-----TCCACACG 951
		: : : : : : : : : : : : : :
Qy	186	RPAC-----C1SSCKPSCCVSSCKPSC--PTTCCTP-----TCHPCTCG 226
		: : : : : : : : : : : : : :
Db	952	TTCATTAAATCTCAATGCG--AACTCTCTCA-TCATTTTCCATTTCCGAATATCTCTCTCT 1008
		: : : : : : : : : : : : : :
Qy	227	SSC 229
		: : : : : : : : : : : : : :
Db	1009	TGC 1011
		: : : : : : : : : : : : : :
RESULT 11		
US-09-627-650B-1		
; Sequence 1, Application US/09627650B		
; Patent No. 6406872		
; GENERAL INFORMATION:		
; APPLICANT: Bamber, Bruce		
; APPLICANT: Jorgensen, Erik		
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and		
; TITLE OF INVENTION: Methods Related Thereto		
; FILE REFERENCE: 21101,000903		
; CURRENT APPLICATION NUMBER: US/09/627,650B		
; CURRENT FILING DATE: 2000-07-28		
; PRIOR APPLICATION NUMBER: 09/436,063		
; PRIOR FILING DATE: 1999-11-08		
; PRIOR APPLICATION NUMBER: 60/107,727		
; PRIOR FILING DATE: 1998-11-09		
; NUMBER OF SEQ ID NOS: 50		
; SOFTWARE: PatentIn Ver. 2.1		
; SEQ ID NO 1		
; LENGTH: 1652		
; TYPE: PRT		
; ORGANISM: Caenorhabditis elegans		
US-09-627-650B-1		

Query Match	21.3%	Score	316.5	DB 4	Length	1652			
Best Local Similarity	25.1%	Pred. No.	1.2e-16						
Matches	76	Conservative	27	Mismatches	111	Indels	89	Gaps	17

Qy	4	SCGSGVCDKUGUGLGG---QESCCRPSGQYTTCCRRITC-CRPSCCGISCCRPSGCISSC	59
		: :	: :
Db	1222	TCCG-CCGTGCGGAATGACATGCAACGGCGGCTCCAAACCTCGATCATCGCGGTATCAG	1280
		: :	: :
Qy	60	CKPSCCLITTCRRITTCRPSGCCISSC-CRPSCCISSC-----CKRPSCC---KPT	103
		: :	: :
Db	1281	CAGTCGAATGATTCGTGTGTCAGTCACAGTCACATTCACATTCGTCACCGGTGCGCGTIT	1340
		: :	: :
Qy	104	CCR-----PSCCISSC-CRPSCCISSCKPSCCRRITTCRPSGCCISSCGRP	147
		: :	: :
Db	1341	CCTCTTGTTTTCATCTTTCAACACTCTCTCTGGCTGATTC-TACTGTACAANITCAA	1399
		: :	: :
Qy	148	SC-CIISGCC-----KPSCCQITTCRPSCCISSCIRPQCQPSGCC	185
		: :	: :
Db	1400	CGGTCTGCCGTATATATAGTGAACACAGAGGTGACCGTTCGGATGC-----TCCACAGCC	1452
		: :	: :
Qy	186	RPAC-----CISSCCHPSCVSSCRCPSC--PTTCCRRITCCTCPHPITCCG	226
		: :	: :
Db	1453	TTCATTATCTCAATCC--AACTTCTCCA-TCATTTTCCATTTCCGAATATCTCTTTTCT	1509
		: :	: :
Qy	227	SSC	229
		: :	: :
Db	1510	TGC	1512
		: :	: :

```

RESULT 12
US-09-436-063C-1
; Sequence 1, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095Ccorrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1652
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-1

Query Match      21.3%  Score 316.5; DB 4; Length 1652;
Best Local Similarity 25.1%  Pred. No. 1.2e-16;
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;

Qy      4  SCCGVCSDQSGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db      1222  TCCG-CCTTGGCGAATGCACATGCAACGGCGCTCCAACCTCGATCATCGCGCTCATCAAG 1280

Qy      60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db      1281  CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGCGCGGCTTT 1340

Qy      104  CCR-----PSCCISSC-CRPSCCISSCCKPSCCCTTCRPSCCISSCCRP 147
Db      1341  CCTCTGTTTTCATCTGTGTTCAACACTCTCTCTGGCTGATTC-TACTGTACAATCAA 1399

Qy      148  SC-CISSCC-----KPSCCOTTCCRPSCCISSCYRPPQCCPSCC 185
Db      1400  GCGTCTGCGGATATTAGTAGAACACGAGGGTGACCGTTGCGATGC-----TCCAGACC 1452

Qy      186  RPAC-----CISSCCHPSCCVSSCRCPSC--PTTCCRT-----TCFHPICCG 226
Db      1453  TTCATTAATCTCAATCC--AACTTCTCTCA-TCATTTTCCATTTCGAATATCTCTTTTCT 1509

Qy      227  SSC 229
Db      1510  TGC 1512

RESULT 13
US-09-627-650B-7
; Sequence 7, Application US/09627650B
; Patent No. 6406872
; GENERAL INFORMATION:
; APPLICANT: Bamber, Bruce
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: 21101.000903
; CURRENT APPLICATION NUMBER: US/09/627,650B
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 09/436,063
; PRIOR FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107,727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508

Qy      4  SCCGVCSDQSGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db      2106  TCCG-CCTTGGCGAATGCACATGCAACGGCGCTCCAACCTCGATCATCGCGCTCATCAAG 2164

Qy      60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db      2165  CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGCGCGGCTTT 2224

Qy      104  CCR-----PSCCISSC-CRPSCCISSCCKPSCCCTTCRPSCCISSCCRP 147
Db      2225  CCTCTGTTTTCATCTGTGTTCAACACTCTCTCTGGCTGATTC-TACTGTACAATCAA 2283

Qy      148  SC-CISSCC-----KPSCCOTTCCRPSCCISSCYRPPQCCPSCC 185
Db      2394  TGC 2396

```

```

; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-627-650B-7

Query Match      21.3%  Score 316.5; DB 4; Length 2508;
Best Local Similarity 25.1%  Pred. No. 1.6e-16;
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;

Qy      4  SCCGVCSDQSGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db      2106  TCCG-CCTTGGCGAATGCACATGCAACGGCGCTCCAACCTCGATCATCGCGCTCATCAAG 2164

Qy      60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db      2165  CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGCGCGGCTTT 2224

Qy      104  CCR-----PSCCISSC-CRPSCCISSCCKPSCCCTTCRPSCCISSCCRP 147
Db      2225  CCTCTGTTTTCATCTGTGTTCAACACTCTCTCTGGCTGATTC-TACTGTACAATCAA 2283

Qy      148  SC-CISSCC-----KPSCCOTTCCRPSCCISSCYRPPQCCPSCC 185
Db      2394  TGC 2396

RESULT 14
US-09-436-063C-7
; Sequence 7, Application US/09436063C
; Patent No. 6407210
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Erik
; TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
; TITLE OF INVENTION: Methods Related Thereto
; FILE REFERENCE: P-1095Ccorrected
; CURRENT APPLICATION NUMBER: US/09/436.063C
; CURRENT FILING DATE: 1999-11-08
; PRIOR APPLICATION NUMBER: 60/107727
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2508
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-436-063C-7

Query Match      21.3%  Score 316.5; DB 4; Length 2508;
Best Local Similarity 25.1%  Pred. No. 1.6e-16;
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;

Qy      4  SCCGVCSDQSGGGLG---QESCCRPSCCQTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db      2106  TCCG-CCTTGGCGAATGCACATGCAACGGCGCTCCAACCTCGATCATCGCGCTCATCAAG 2164

Qy      60  CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISSC-----CCRPSCC---RTT 103
Db      2165  CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTCGATCGTCAACCGCGCGGCTTT 2224

Qy      104  CCR-----PSCCISSC-CRPSCCISSCCKPSCCCTTCRPSCCISSCCRP 147
Db      2225  CCTCTGTTTTCATCTGTGTTCAACACTCTCTCTGGCTGATTC-TACTGTACAATCAA 2283

Qy      148  SC-CISSCC-----KPSCCOTTCCRPSCCISSCYRPPQCCPSCC 185

```

1b 2284 GCCTCTGCCGTATATTAGTGAACACAGAGGGTGACCGTTGCGATGC-----TCCAGACC 2336
QY 186 RPAC-----CISSCCHPSCVSSCRCPFSC--PTTCRT-----TCFHPICCG 226
Db 2337 TTCATTAATCTCAATCC--AACTTCTCA-TCAATTTCCATTTCGAATATCTCTTTTCT 2393
QY 227 SSC 229
Db 2394 TGC 2396

RESULT 15

US-09-627-650B-3
: Sequence 3, Application US/09627650B
: Patent No. 6406872
: GENERAL INFORMATION:
: APPLICANT: Bamber, Bruce
: APPLICANT: Jorgensen, Erik
: TITLE OF INVENTION: Nematode Neuromuscular Junction GABA Receptors and
: TITLE OF INVENTION: Methods Related Thereto
: FILE REFERENCE: 21101.000903
: CURRENT APPLICATION NUMBER: US/09/627,650B
: CURRENT FILING DATE: 2000-07-28
: PRIOR APPLICATION NUMBER: 09/436,063
: PRIOR FILING DATE: 1999-11-08
: PRIOR APPLICATION NUMBER: 60/107,727
: PRIOR FILING DATE: 1998-11-09
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 2544
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans
US-09-627-650B-3

Query Match 21.3%; Score 316.5; DB 4; Length 2544;
Best Local Similarity 25.1%; Pred. No. 1.6e-16;
Matches 76; Conservative 27; Mismatches 111; Indels 89; Gaps 17;
QY 4 SCCGVCSDSCGGGLG---QESCRPSCGTTCCRTTC-CRPSCCISSCCRPSCCISSC 59
Db 2142 TCGC-CGTTGCCGAATGCACATGCAGCGGCTCCAACCTCGATCATCGCCGTCATCAAG 2200
QY 60 CKPSCCLTTCRTTCRPSCCISSC-CRPSCCISS-----CKKPSCC---RTT 103
Db 2201 CAGTCGAATCGATTCTGTGTCAGTCACAGTCACATTGACATCGTCAGCGCGCGGTTT 2260
QY 104 CCR-----PSCCISSC-CRPSCCISSCKPSCRTTCRPSCCISSCCRP 147
Db 2261 CCTCTTTGTTTCATCTTGTTCAACACTCTCTTGGCTGATTC-TACTGTACAAATCCAA 2319
QY 148 SC-CISSC-----KPSCGTTCCRPSCCISSCYRPOCCQPSCC 185
Db 2320 CGGTCTGCCGTATATTAGTGAACACAGGGTGACCGTTGCCATGC-----TCCAGACC 2372
QY 186 RPAC-----CISSCCHPSCVSSCRCPFSC--PTTCRT-----TCFHPICCG 226
Db 2373 TTCATTAATCTCAATCC--AACTTCTCA-TCAATTTCCATTTCGAATATCTCTTTTCT 2429
QY 227 SSC 229
Db 2430 TGC 2432

Search completed: June 11, 2003, 11:06:58
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 11:02:46 : Search time 78 Seconds
(without alignments)
607.575 Million cell updates/sec

Title: US-09-874-062-3
Perfect score: 1484
Sequence: 1 MVSSCGSVCSQSCGGGLG.....PTTCRTTCFHPICGSSCC 230

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_protist.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacterioph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	996.5	67.1	195	11 Q9D0X9	Q9d0x9 mus musculus
2	968	65.2	210	4 Q9BYR0	Q9byr0 homo sapien
3	925	62.3	201	4 Q9BQ66	Q9bq66 homo sapien
4	898	60.5	195	4 Q9BYQ6	Q9byq6 homo sapien
5	866.5	58.4	191	4 Q9BYQ8	Q9byq8 homo sapien
6	858.5	57.9	193	4 Q9BYQ5	Q9byq5 homo sapien
7	808.5	54.5	202	11 Q91W93	Q91w93 mus musculus
8	806	54.3	165	11 Q9D7P3	Q9d7p3 mus musculus
9	804.5	54.2	186	4 Q9BYR2	Q9byr2 homo sapien
10	800.5	53.9	166	4 Q9BYR3	Q9byr3 homo sapien
11	764.5	51.5	168	11 Q9D732	Q9d732 mus musculus
12	744	50.1	195	11 Q9D141	Q9d141 mus musculus
13	738	49.7	167	11 Q9D122	Q9d122 mus musculus
14	709	47.8	191	11 Q9D3H7	Q9d3h7 mus musculus
15	701.5	47.3	186	11 Q64526	Q64526 mus musculus
16	693.5	46.7	223	11 Q62220	Q62220 mus musculus

17	684	46.1	230	11 Q64507	Q64507 mus musculus
18	654.5	44.1	136	4 Q9BYR5	Q9byr5 homo sapien
19	645	43.5	98	4 Q9BYR4	Q9byr4 homo sapien
20	619	41.7	151	6 Q28576	Q28576 ovis aries
21	584.5	39.4	127	4 Q9BYQ7	Q9byq7 homo sapien
22	584.5	39.4	174	4 Q9BYQ4	Q9byq4 homo sapien
23	569.5	38.4	126	6 Q28707	Q28707 oryctolagus
24	560	37.7	188	11 Q70148	Q70148 rattus norv
25	546	36.8	227	11 Q9D225	Q9d225 mus musculus
26	545.5	36.8	159	4 Q9BYQ0	Q9byq0 homo sapien
27	542.5	36.6	159	4 Q9BYQ3	Q9byq3 homo sapien
28	530	35.7	154	4 Q9BYQ2	Q9byq2 homo sapien
29	521	35.1	114	4 Q9BYQ9	Q9byq9 homo sapien
30	515	34.7	154	4 Q9BYP9	Q9byp9 homo sapien
31	481.5	32.4	106	4 Q9BYU7	Q9byu7 homo sapien
32	463.5	31.2	175	4 Q07628	Q07628 homo sapien
33	451	30.4	177	4 Q07627	Q07627 homo sapien
34	447	30.1	174	4 Q9BYS1	Q9bys1 homo sapien
35	438.5	29.5	167	4 Q9BYS2	Q9bys2 homo sapien
36	432	29.1	182	6 Q29619	Q29619 ovis aries
37	431.5	29.1	169	4 Q14564	Q14564 homo sapien
38	428.5	28.9	189	11 Q9D527	Q9d527 mus musculus
39	421.5	28.4	177	11 Q9D644	Q9d644 mus musculus
40	380	25.6	130	11 Q9Z287	Q9z287 mus musculus
41	376.5	25.4	111	4 Q9BYQ1	Q9byq1 homo sapien
42	375	25.3	661	5 Q17982	Q17982 caenorhabdi
43	366.5	24.7	136	11 Q9D718	Q9d718 mus musculus
44	360.5	24.3	152	6 Q29620	Q29620 ovis aries
45	358	24.1	251	5 Q24774	Q24774 enclytraeus

ALIGNMENTS

RESULT 1

Q9D0X9	PRELIMINARY;	PRT: 195 AA.
ID Q9D0X9		
AC Q9D0X9		
DT 01-JUN-2001 (TREMBLrel. 17, Created)		
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE I110054P19rik protein.		
GN I110054P19rik.		
OS Mus musculus (Mouse).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX NCBI_TaxID=10090;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;		
RX MEDLINE=21085660; PubMed=11217851;		
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,		
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,		
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA Nordone P., Ring B., Kingwald M., Rodriguez I., Sakamoto N.,		
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Whiling L.,		
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,		
RA Hayashizaki Y.;		
RT "Functional annotation of a full-length mouse cDNA collection.";		
RL Nature 409:685-690(2001).		
DR EMBL; AK004258; UAB23238.1; ..		
HSSP; 046655; 1CJH.		

[illegible]

Page 3

Db	106300	TTGAATCTCAAAATTCAGTCAAAATCCACATCACAATGTTTTAGAACCTCTTTATTTCTAATT	106330
QY	961	CAATATACATAAATCTTCAAAATGGTATCCTTCTAGATGCTTTCTTCCCTCAATGTTTCTGTG	1020
Db	106350	CAATATACATAAATCTTCAAAATGGTATCCTTCTAGATGCTTTCTTCCCTCAATGTTTCTGTG	106411
QY	1021	GTATCAATTTTTCATGTTGGAAATGTTTGTGATGTTCTCAATAAATTCATAGTGTCAAAA	1080
Db	106420	GTATCAATTTTTCATGTTGGAAATGTTTGTGATGTTCTCAATAAATTCATAGTGTCAAAA	106471
QY	1081	GCACAAAAA	1090
Db	106480	GCACAAAAA	106489

LOCUS	AC006070	161987 bp	DNA	linear	PRI 17-DEC-1998							
DEFINITION	Homo sapiens chromosome 17, clone hRPK.206_C20, complete sequence.											
ACCESSION	AC006070											
VERSION	AC006070.1	GI:4028942										
KEYWORDS	HTG.											
SOURCE	Homo sapiens.											
ORGANISM	Homo sapiens											
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;											
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.											
REFERENCE	1 (bases 1 to 161987)											
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.											
TITLE	Homo sapiens chromosome 17, clone hRPK.206_C20											
JOURNAL	Unpublished											
REFERENCE	2 (bases 1 to 161987)											
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Boutwell, C., Brown, A., Castle, A., Cerny, J., Collangelo, M., Collins, S., Collymore, A., Cooke, P., Corliss, D., Depayre, E., Devon, K., Dewar, K., Donelan, L., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S., Geraghty, K., Grant, G., Hagos, B., Heaford, A., Herena, L., Horton, L., Howland, J. C., Jacotot, L., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Nahf, R., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Tichovolsky, N., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D., Ye, W. J., Zhao, J. and Zody, M.											
TITLE	Direct Submission											
JOURNAL	Submitted (26-NOV-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA											
REFERENCE	3 (bases 1 to 161987)											
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Depayre, E., Devon, K., Dewar, K., Donelan, L., Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Herena, L., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.											
TITLE	Direct Submission											
JOURNAL	Submitted (17-DEC-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA											
COMMENT	On Dec 17, 1998 this sequence version replaced gi:4006836. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html											

Only the last 161987 base pairs of this clone are being submitted

Only the last 161987 base pairs of this clone are being submitted.

AC025904
AC025904.2. GI:7382590
HTG: HTGS_PHASE1, HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE
Homo sapiens, clone RP11-585G19
JOURNAL
Unpublished
REFERENCE
1. (bases 1 to 174032)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beka, F., Boguski, M., Bouck, G., Brown, A., Burkett, G., Campione, A., Castle, A., Choquel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Houton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczkv, J., Levine, R., Li, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., Mihova, T., McKernan, K., McPheeters, R., Meidrid, J., Meneses, L., Mihov, T., Miranda, C., Miñana, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zhan, A. and Zody, M.
Direct Submission
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 1, 2000 this sequence version replaced gi:7249428.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research.
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8441
Center clone name: 585_G_19
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158771 bases at least Q40
Consensus quality: 167389 bases at least Q30
Consensus quality: 170660 bases at least Q20
Insert size: 170000; agarose-ff
Insert size: 172432; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-ff
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2588: contig of 2588 bp in length
* 2589 2688: gap of 100 bp
* 2689 6465: contig of 3777 bp in length
* 6466 6565: gap of 100 bp
* 6566 8846: contig of 2281 bp in length

GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 11:02:46 ; Search time 18 Seconds
(without alignments)
1228.384 Million cell updates/sec

Title: US-09-874-062-3

Perfect score: 1484

Sequence: 1 MVSSCGSVCSQSGGGLG.....PTTCRTTCFHPICGGSCC 230

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	701.5	47.3	186	A45910	ultra-high-sulfur
2	693.5	46.7	223	B38346	ultra-high-sulfur
3	684	46.1	230	A38346	ultra-high-sulfur
4	619	41.7	151	S60314	hair keratin cyste
5	569.5	38.4	126	I46489	cysteine-rich hair
6	560	37.7	188	JC6547	high-sulfur protei
7	463.5	31.2	175	S37649	high-sulfur kerati
8	451	30.4	177	S37650	high-sulfur kerati
9	434	29.2	182	KRSHHD	keratin high-sulfu
10	432	29.1	182	I47105	high-sulfur wool m
11	431.5	29.1	169	S18946	ultra high-sulfur
12	408.5	27.5	172	KRSHHA	keratin high-sulfu
13	402	27.1	172	I47106	high-sulfur wool m
14	384.5	25.9	162	I47107	high-sulfur wool m
15	375	25.3	654	T30136	hypothetical prote
16	360.5	24.3	152	I47108	keratin high-sulfu
17	358.5	24.2	152	KRSHHC	keratin high-sulfu
18	358.5	24.2	152	I47111	high-sulfur wool m
19	358	24.1	251	A55035	cysteine-rich prot
20	350.5	23.6	152	I47109	high-sulfur wool m
21	350.5	23.6	152	I47112	high-sulfur wool m
22	348	23.5	156	KRSHHB	keratin high-sulfu
23	344	23.2	191	I46412	keratin KAP5.4 - s
24	325	21.9	197	I46413	keratin KAP5.5 - s
25	305.5	20.6	182	A36686	ultra-high-sulfur
26	298.5	20.1	572	T29880	hypothetical prote
27	296.5	20.0	131	KRGT3M	keratin high-sulfu
28	294	19.8	728	T20561	hypothetical prote
29	293	19.7	425	T18592	hypothetical prote

ALIGNMENTS

RESULT 1

A45910

ultra-high-sulfur keratin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 24-Sep-1999

C:Accession: A45910

R:McNab, A.R.; Wood, L.; Theriault, N.; Gierman, T.; Vogeli, G.

J. Invest. Dermatol. 92, 263-266, 1989

A:Title: An ultra-high sulfur keratin gene is expressed specifically during hair growth

A:Reference number: A45910; MUID:89140394; PMID:2465353

A:Accession: A45910

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-186 <MCN>

A:Cross-references: GB:M27685; MID:g341749; PIDN:AAA81560.1; PID:g1066818

C:Superfamily: ultra-high-sulfur keratin

Query Match 47.3%; Score 701.5; DB 2; Length 186;

Best Local Similarity 49.0%; Pred. No. 1.le-42;

Matches 119; Conservative 20; Mismatches 25; Indels 75; Gaps 11;

QY	1	MVSSCGSVCSQSGGGLGQESCRPSCCOTTCRTTCRRPSC---CISCCRPSCCIS 57
DB	1	MTNSCC-----SPCQPTCCRTTCRTTCRRPSCVTSQCCQPSCCGS 45
QY	58	SCCKPSCCLTTCRTTCRRPSCCISCCRPSCCISCCRPSCCCTTCRRPSCCISCCRP 117
DB	46	SCCQP-CCQTTCCRT-----CFQPCCVSSCCRTTCCQPCCVSSCCQP 87
QY	118	SCCISCCRPSCCCTTCRRPSCCISCCRPSCCISCCRPSCCCTTCRRPSCCISCCRP 177
DB	88	-----CCQSCCSCCQPCRCESCCQPCRCISCCQP-----CCRPSCCSCCQP 135
QY	178	OCQPCRCACCTTCRRPSCCISCCRPSCCISCCRPSCCCTTCRTTCRRPSCCISCCRP 227
DB	136	CCQP-----FCLNCCQPCAC-----SGPVTCTR-TCYQTCVCVPCCLSGQCGS 178
QY	228	SCC 230
DB	179	SCC 181

RESULT 2

B38346

ultra-high-sulfur keratin 2 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999

C:Accession: B38660; B38346

R:Wood, L.; Mills, M.; Hatzenbuehler, N.; Vogeli, G.

J. Biol. Chem. 266, 4024, 1991

A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin

A:Reference number: A38660; MUID:91154184; PMID:1840598